SENTENTIAL CROSS-SITUATIONAL LEARNING: THE COMBINED EFFORTS OF STATISTICAL AND SYNTACTIC CUES IN WORD LEARNING

Examination Number: B004504

A dissertation submitted in partial fulfillment of the requirements for the degree of

Master of Science in Evolution of Language and Cognition

The University of Edinburgh

August 2011

Supervised by: Dr. Kenny Smith

ABSTRACT

Recently, cross-situational word learning (XSL) has gained attention as a viable mechanism for learning the meaning of words. XSL refers to the tracking of word-referent pairs over a series of different exposures (Siskind, 1996; Yu & Smith, 2007). However, it has recently been suggested that word learning likely functions by the extraction of information from several cues, such as perceptual statistical, social, and linguistic (Hirsh-Pasek, Golinkoff, & Hollich, 2000). In this work, I measured the contribution of co-occurrence frequencies (statistical cues) and syntactic-semantic (linguistic cues) links in a sentential XSL experiment that contained both nouns and verbs. Participants were exposed to three different learning conditions that varied in the mode of presentation. Participants either learned nouns and verbs together in one phase (unstaggered learning), or learned in two phases, initially learning the subset of nouns (staggered-with-nouns) or verbs (staggered-with-verbs) first. With unstaggered learning, participants could use primarily only XSL, while with staggered learning, participants could use both XSL and syntactic bootstrapping. Results demonstrated that the combination of syntactic bootstrapping and XSL significantly facilitated word learning. The results support an integrative and comprehensive model of word learning (e.g. Hirsh-Pasek et al., 2000). In the general discussion, the relationship between an integrative view of language and a domain-general account of language is discussed.

ABBREVIATIONS

C - Level of referential uncertainty in XSL studies

HSP – Human Simulation Paradigm

MD – Mean Difference

POMP – Perceptual or memory primitive

SL – Statistical Learning

STAGNOUN – Staggered learning condition where nouns are learned first

STAGVERB – Staggered learning condition where verbs are learned first

UNSTAG – Unstaggered learning condition where nouns and verbs are learned simultaneously

XSL – Cross-situational learning

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Language acquisition research aims at discovering the mechanisms children normally use to learn language with such ease and rapidity (Cattell, 2000). One of the fundamental steps in language acquisition, which children achieve at a very young age, is word learning (Bloom, 2000). Theoretically, word learning should pose a difficult problem due to the enormous amount of referential uncertainty in the linguistic environment (e.g. Quine, 1960). In Quine's (1960) famous example of referential uncertainty, a field linguist hears native speakers of an unknown language say "gavagai" just as a rabbit runs by and he attempts to understand the meaning of the word. The linguist is left with numerous possible referents for "gavagai", such as "rabbit", but certainly also "legs" or "the colour brown", "the action of running", "possible food", or "time for hunting". For children acquiring their native language, there are an infinite number of possible word-to-referent mappings (Yu & Smith, 2007). Despite this referential uncertainty, they normally acquire the lexicon and grammar of their native language rapidly and effortlessly. Children are speaking over 200 words by 21 months of age (Fenson, Dale, Reznick, Bates, & Thal, 1994) and around 60,000 by 18 years of age (Bloom, 2000).

Accounts of word learning have focused on a variety of cognitive heuristics (e.g. Markman, 1989), social learning (Tomasello, 2000), linguistic factors (Gleitman, 1990; Pinker, 1989), or on a cross-situational learning (henceforth XSL) approach (Siskind, 1996; Smith, Smith, & Blythe, 2010, Yu & Smith, 2007).

In the present work, I will present an XSL learning experiment of two-word sentences that incorporates both statistical and linguistic information. The experiment demonstrates how learners perform better when they extract both statistical and syntactic cues perform better than when they extract only statistical cues. This experiment will be used to support an integrative model of word learning that combines several individual factors into a comprehensive account (e.g. Hirsh-Pasek, Gollinkoff, & Hollich, 2000).

This work is broken down into six chapters. Chapter 2 provides a background account of four different perspectives on word learning and on an integrative theory of word learning known as the emergentist coalition (Hirsh-Pasek et al., 2000). This idea is then generalized to other aspects of language acquisition based on findings from SL research. Chapter 3 focuses on the two mechanisms of word learning, XSL and bootstrapping, and discusses studies that examine the effect of the combined efforts of statistical and syntactic information on word learning. Chapter 4 presents an original research experiment that further explores the combination of XSL and syntactic bootstrapping during a word learning experiment. XSL is a statistical cue that refers to sensitivity to the frequencies of co-occurrences between a word and a possible referent (Siskind, 1996; Smith et al., 2010; Yu & Smith, 2010). Syntactic bootstrapping refers to the use of the syntactic structure of a sentence to aid lexical knowledge (Gleitman, 1990). The experiment examines the possibility of simultaneous noun and verb learning, referred to as sentential XSL, while testing for the effect of syntactic bootstrapping and bootstrapping through prior knowledge. The results, which reveal that the incorporation of bootstrapping significantly increased accuracy, indicate that the combination statistical and linguistic cues facilitated word learning more than just the use of statistical cues alone and offer support for an integrative model of word learning (Hirsh-Pasek et al., 2000; Hollich et al., 2000). Chapter 5 discusses the issues that an integrative account of language acquisition brings to a domain-general account of language acquisition. Finally, chapter 6 briefly summarizes the entire dissertation and provides concluding remarks.

CHAPTER 2: ACCOUNTS OF WORD LERANING AND LANGUAGE ACQUISITION

2.1. ACCOUNTS OF WORD LEARNING

There are several theories that examine lexical development from different angles, such as the constraints theory of word learning (e.g. Behrend, Scofield, & Kleinknecht, 2001), the social-pragmatic theory (Tomasello, 2000), the two bootstrapping theories (e.g. Gleitman, 1990) and the XSL theory. The constraints theory contends that children are innately equipped with several word-learning biases or heuristics that help them learn the meaning of a word in one or few exposures, a phenomenon known as "fast-mapping" (Behrend et al., 2001). Included in this account are the whole object bias, (Macnamara, 1982), the mutual exclusivity bias (Markman, 1989), and the shape bias (Landau, Smith, & Jones, 1998). The social-pragmatic theory argues that cultural learning, which includes factors such as joint attention between a child and adult (Tomasello, 1999) and the understanding of communicative intents (Behne, Carpenter, & Tomasello, 2005), is crucial for word learning (Tomasello, 2000). Related to the social-pragmatic theory is the idea that children use their theory of mind to determine the object to which a person is referring (Bloom, 2000). The syntactic and semantic bootstrapping hypotheses (Gleitman, 1990; Pinker, 1989) state that children easily learn language because they use their initial acquired knowledge as "bootstraps" to acquire further (Pinker, 1989; Gleitman, 1990). In particular, the linguistic cues of syntactic-semantics mappings can be used as bootstraps to facilitate lexical development (Bloom, 1990, 1994; Gleitman, 1990). For example, it has been claimed that the syntactic structure of verbs is a projection of their semantic structure (Fisher, Gleitman, & Gleitman, 1991). Mental verbs, such as *think*, *feel*, and *believe*, all take a sentence complement. Bootstrapping theorists argue that children can indeed notice these links and use them to bootstrap lexical development. The two main theories of bootstrapping are syntactic bootstrapping (Gleitman, 1990) and semantic bootstrapping (Pinker, 1984; 1989). The former argues that children use knowledge of syntactic structure to bootstrap word learning (Gleitman, 1990), while the latter is the reverse and argues that children use knowledge of semantics to bootstrap knowledge of syntax (Pinker, 1984). Finally, the XSL theory argues that learners are sensitive to the co-occurrence frequencies between a word and its possible referents over a series of exposures (Yu & Smith, 2007). There is a growing body of evidence that suggests that adults (Smith, Smith & Blythe, 2010; Yu & Smith, 2007) and infants (Smith & Yu, 2008) can detect cross-situational statistics. XSL is a type of statistical learning (henceforth SL) that is categorized as a "first-order" statistic, relative to higher-order statistics like conditional probabilities (Aslin & Newport, 2008; Fiser & Aslin, 2002). Prior to the recent onset of XSL experiments, researchers often referred to this idea as a "simple" or "dumb" associative learning mechanism (Smith, et al., 1996). Further information regarding XSL and the mechanism behind it will be discussed in Chapter 3.

Language acquisition is often viewed as occurring based primarily on one of three perspectives (i.e. word learning heuristics, social-pragmatics, or SL) (Hirsh-Pasek, Golinkoff, Hennon, & McGuire, 2004). In such cases, work on one aspect of word learning can offer at most only a partial understanding of word learning. For example, a constraints model of word learning does not explain why 18-month old children benefit from social cues but 10-month old children do not (Pruden, Hirsh-Pasek, Golinkoff & Hennon, 2006), and an XSL model of word learning does not explain how children fast-map the meaning of a word (Behrend et al., 2001). An idea that is gaining popularity is that learners use *multiple* cues that interact with each other to develop their mental lexicon (e.g. Frank, Goodman, & Tenenbaum, 2007; Smith et al., *under revision*). For example, Smith et al. (*under revision*) discuss the effectiveness of XSL after the use of word-learning heuristics to constrain the degree of referential uncertainty. Frank et al. (2007) introduce a Bayesian model for word learning that incorporates both co-occurrence frequencies and social cues. Moreover, several authors have collaborated to create the emergentist coalition model of word learning (e.g. Hollich et al., 2000; Hirsh-Pasek et al., 2004). The emergentist coalition model states that children are biased to be sensitive to multiple cues, including statistical, social, attentional, cognitive, and linguistic. Furthermore, it states that the reliance on the cues shifts with development, so that attentional and statistical cues are important during the initial stages of word learning while social cues are more important later on. Smith et al. (*under revision*) and Frank et al.'s (2007) discussions and the emergentist coalition model all emphasize the importance of multiple cues in word learning, based on statistics, heuristics, linguistics, or pragmatics.

2.2. TOWARD AN INTEGRATIVE ACCOUNT OF LANGUAGE ACQUISITION

The idea that multiple cues mutually reinforce one another is not specific to word learning, but can apply to other aspects of language acquisition (e.g. Romberg & Saffran, 2010). One area of language acquisition research that is embracing the idea of multiple correlating cues facilitating language learning is SL. SL has been defined as the "acquisition of structured information from the auditory or visual environment via sensitivity to frequency or probability distributions" (Aslin & Newport, 2008). Based on mere exposure to the linguistic input, children can acquire the regularities, often subconsciously, and predict future events. Natural languages consist of regular surface distributional patterns that learners can extract and interpret. SL proponents argue that categorization mechanisms are used for learning speech categories (e.g. Maye, Werker, & Gerken, 2002); frequencies and sequential transitional probabilities for segmenting words (Saffran, Aslin, & Newport, 1996); conditional probabilities for learning hierarchical phrase structure (Gómez & Gerken, 1999; Saffran, 2002), and co-occurrence frequencies for learning words (Smith et al., 2010; Yu & Smith, 2007).

A growing body of research demonstrates that multiple, correlated cues in SL significantly facilitate language learning (e.g. Christiansen & Curtin, 2005; Cunillera et al., 2010a; Gerken, Wilson, & Lewis, 2005; Newport & Aslin, 2004; Sahni et al., 2010). The cues can be the *statistical*, such as transitional probabilities (Saffran et al., 1996), distributional cues (Gerken et al., 2005), and co-occurrence frequencies (Yu & Smith, 2007); *linguistic*, such as the syntactic frame of a sentence, including classifiers or modifiers (Yoshida & Smith, 2005), distributional cues, including bigrams (Monaghan & Christiansen, 2008), syntactic-semantic links that provide an opportunity for bootstrapping (Gilllette et al., 1999), phonotactics (Mattys & Jusczyk, 2001) and lexical stress (Thiesson & Saffran, 2003); *nonlinguistic,* such as social cues (Frank, Goodman, & Tenenbaum, 2009) and prior knowledge (Yurovsky et al., 2010); or *nonlinguistic and multimodal*, such as visual cues (Cunillera, Càmara, Laine, & Rodríguez-Fornells, 2010b).

In summary, the idea that learners can extract multiple correlating cues can be applied both specifically to word learning (Hirsh-Pasek et al., 2000) and more generally to statistical learning of language (Christiansen & Curtin, 2005; Cunillera, Càmara, Laine, & Rodríguez-Fornells, 2010a; Romberg & Saffran, 2000). Chapter 3 of this work focuses on two types of cues that facilitate word learning: co-occurrence frequencies, which are statistical cues, and the syntactic-semantic links, which are linguistic cues upon which bootstrapping functions (e.g. Gleitman, 1990).

CHAPTER 3: COMBINING STATISTICAL AND SYNTACTIC CUES IN WORD LEARNING

3.1. PREVIOUS XSL EXPERIMENTS

There are several different experimental paradigms that test for XSL, including the Human Simulation Paradigm (e.g. Gillette et al., 1999), and the Yu and Smith paradigm (Yu & Smith, 2007), and the Smith et al. paradigm (Smith, Smith, & Blythe, 2010). These paradigms differ in stimulus materials, in training and testing methods, and in the type of words (e.g. noun or verb) they examine.

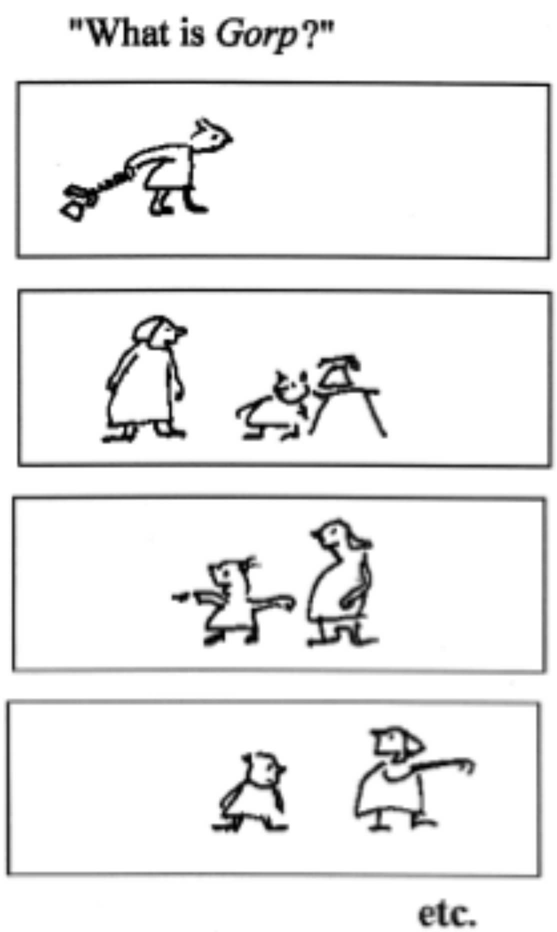
3.1.1 Human Simulation Paradigm and XSL of verbs

One of the early empirical tests for XSL of nouns and verbs in adults is known as the Human Simulation Paradigm (henceforth HSP), designed by Gillette and colleagues (1999). This experimental design uses video samples of natural stimuli with little or no audio input. In Gillette et al.'s experiment, the participants were tested on their ability to identify the 24 most frequent nouns and 24 most frequent verbs. The stimuli were taken from a video sample of child-directed speech between a mother and an 18-24 month old child. There was no audio input except for a beep that indicated the place of the target word. Participants were exposed to the word a total of 6 times. The HSP used a repeated testing approach, where participants were tested on their knowledge of the word after each exposure. Analysis of the final trial revealed that 45% of the nouns and only 15% of the verbs could be identified by all participants when using only cross-situational observation. A subset of verbs particularly difficult for identification was mental verbs (e.g. *think*, *know*). However, a trial-by-trial analysis revealed that identification of both nouns and verbs improved throughout trials, which Gillette et al. (1999) interpreted as the functioning of XSL.

A second experiment revealed imageability, the degree to which a word can be concretely visualized, to be a key predictor for success in identifying verbs but not nouns. In this experiment, verbs with low imageability were more difficult to learn than verbs with high imageability. Gillette et al. concluded that, because verbs are less imageable than nouns, they are more difficult to learn than nouns. Other experiments using the HSP indicate that both adults (Snedeker, Gleitman, & Brent, 1999) and children (Piccin & Waxman, 2007) have more difficulty learning verbs than nouns.

A third experiment performed by Gillette et al., demonstrated how *syntactic* information could be combined with statistical information to improve accuracy in verb learning. This experiment had several different conditions with varying amounts of linguistic information. Statistical information was given in the form of videos where XSL was possible, and linguistic information was given in the form of the sentences which identified the syntactic frame surrounding the target word. The sentences contained a nonsense word to replace the target word, and all other nouns as either the actual nouns or as nonsense nouns. Figure 3.2 reproduces diagrams of three of the six conditions used in the HSP experiment.

a)



b)

Why don't ver GORP telfa?
GORP wastorn, GORP wastorn.
Verg gonna GORP waston?
Mek gonna GORP litch.

etc.

c)

Why don't you GORP Grandma?
GORP Daddy, GORP Daddy.
You gonna GORP Daddy?
I'm gonna GORP Markie.

etc.

Figure 3.1: Examples of three of six conditions taken from Gillette et al's (1999) HSP. *A* is a cartoon approximation of four of the six videos used in the condition, where there was statistical but not linguistic information. *B* is an excerpt from the sentences used in the condition, where there was statistical and linguistic information, but where the nouns were replaced by nonsense verbs. *C* is an excerpt from the sentences used in the condition where there was statistical and linguistic information. All stimuli are reproduced from Gillette et al. (1999).

In the condition where there was only linguistic information and no statistical information, participants were able to identify over 50% of the verbs. In the condition where there was both statistical and linguistic information, but with all nouns replaced by nonsense words, participants identified 75% of the verbs. Finally, in the condition where there was both statistical and full linguistic information, participants were able to identify 90% of the target verbs. Gillette et al. concluded that linguistic context is very important for learning verbs and particularly for the subset of verbs with a low degree of imageability.

3.1.2. Yu and Smith paradigm

The Yu and Smith paradigm of XSL, named after its designers, Chen Yu and Linda Smith (Yu & Smith, 2007), allows for controlled levels of referential uncertainty in the laboratory. In Yu and Smith's experiment, participants simultaneously saw pictures of uncommon objects (e.g. canister, rasp, and facial sauna) and heard and saw pseudowords which referred to the objects. Participants learned the names for objects using 2x2, 3x3, and 4x4 conditions in the training phases. The 2x2 condition displayed 2 labels and 2 possible referents; the 3x3 condition displayed 3 labels and 3 possible referents; and the 4x4 condition displayed 4 label and 4 possible conditions. This yielded a total number of 4, 9, and 16 potential word-referent associations, respectively, per trial. Each individual trial contained multiple labels and multiple referents but each label had one referent it consistently co-occurred with in each trial. Figure 3.1, below, is a schematic illustration of Yu and Smith's 3x3 condition.

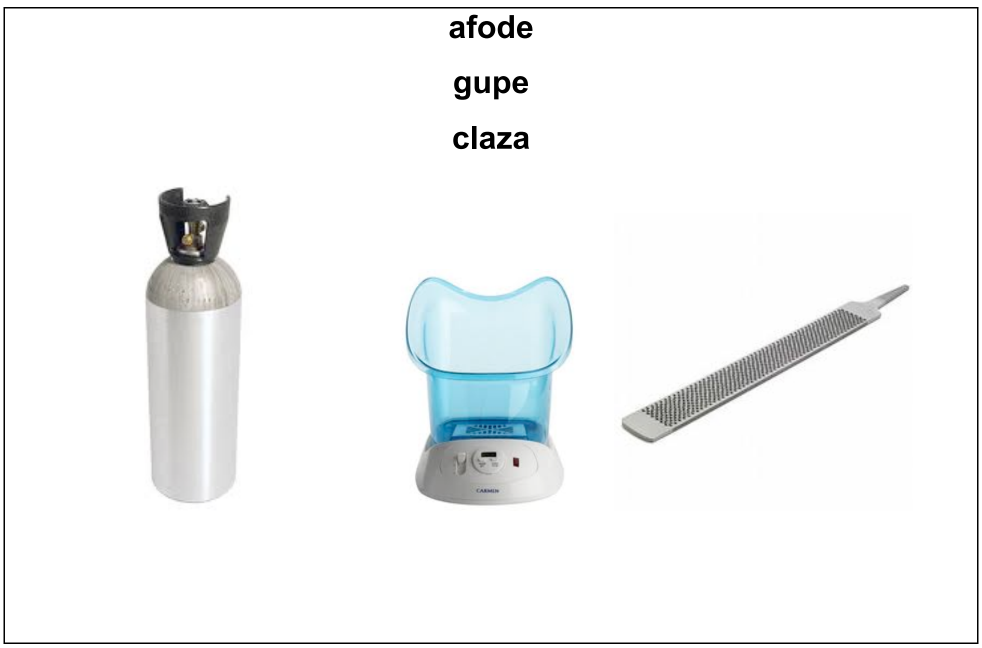


Figure 3.2: An approximation of Yu and Smith's (2007) 3x3 condition, using a gas canister, facial sauna, and a rasp. N.B: This figure is not affiliated with the authors themselves. Although these are objects that the authors used, these are not necessarily the pictures of the objects that were used during Yu and Smith's experiment. The pseudowords were also not necessarily used in Yu and Smith's experiment.

In the experiment, there was within-trial ambiguity, but between-trial consistency. Word learning was possible over multiple trials. In order to succeed, participants had to keep track of multiple word-referent pairs, eliminate misleading word-referent pairs, and select the correct pair. Participants were instructed to learn 18 words in each condition. Each word-referent pair occurred a total of 6 times. The length of the trials varied per condition. The trial length was 6 seconds in the 2x2 condition, 9 seconds in all 3x3 condition, and 12 seconds in the 4x4 trial, which yielded a total of 324 seconds for each condition. In the test phase, participants were shown 1 label and 4 pictures and were asked to select the picture they thought corresponded to the label given to them. Yu and Smith found that, on average, participants performed significantly better than chance with all three degrees of referential uncertainty. Adults successfully learned over 16 of 18 pairs in the 2x2 condition, more than 13 of 18 pairs in the 3x3 condition, and almost 10 of 18 pairs in the 4x4 condition. Based on their results, they concluded that adults are able to rapidly learn in a cross-situational manner even when faced with high referential uncertainty, but that accuracy decreased with the increase of referential uncertainty. Smith and Yu (2008) also found that infants as young as 12- and 14-months were sensitive to co-occurrence statistics.

Yu and Smith were also interested in determining the relationship between the incorrect choice of a foil in the test phase and the probability of a spurious correlation involving the foil in the training phase. A spurious correlation is a seemingly correct but misleading correlation between a word and an incorrect referent. Using only the 4x4 condition, they ran a second experiment where they varied the total number of words to be learned and the number of repetitions of each pair. This experiment had three conditions. The first condition consisted of 9 words with 8 repetitions each; the second consisted of 9 words with 12 repetitions each, and the third consisted of 18 words with 6 repetitions each. Here, they found that participants learned more pairs in the 18-word condition than in the 9-word condition, even though they were confronted with twice the number of words. These results indicate that total number of spurious correlations may also come into play during XSL. A higher number of spurious correlations results in higher foil probabilities, which are probabilities between a word and an incorrect referent. If there are fewer words to be learned, a word may frequently appear with an incorrect referent and cause learners to mistake it as a correct pair. If there are more words to be learned, there are more referents for a word to pair with and therefore lower foil probabilities. However, it should be noted that there was no significant difference in performance, based on accuracy proportion, in any of the three conditions.

Scott (2010) adopted the Yu and Smith paradigm to study XSL of verbs in young children. In Scott's study, 31-month old children were tested on XSL of four intransitive (e.g. "She's pimming") and four transitive (e.g. "She's pimming the ball") verbs. The children were separated into high and low English vocabulary groups. The experiment used a 2x2 design and a preferential-looking paradigm to test for accuracy. Scott found that the children in both the high and low vocabulary groups were able to successfully learn the referents for all novel intransitive verbs, as children looked significantly more at the targets rather than distractors. However, for transitive verbs, there was a difference between the high- and low-vocabulary groups. The high vocabulary group was able to successfully learn the transitive verbs, but the low vocabulary group was not. In her discussion, Scott mentions that the discrepancy between the two groups may possibly be due to limitations on working memory, as recent work reveals that vocabulary size is correlated with working memory (Marchman & Fernald, 2008). Transitive verbs have the object of the sentence as an additional argument that children must keep track of, which may induce a greater load on working memory or on general language processing mechanisms (Scott, 2010).

Smith, Smith, and Blythe (*under revision*) identified a flaw in the Yu and Smith paradigm. Yu and Smith (2007) consistently used a 1x4 condition in their test phase; each test thus consisted of 1 label and 4 possible referents. However, the total numbers of referents was 18 for all trials because there were 18 words being learned in each condition. The 1x4 testing design lowered the actual degree of referential uncertainty in all conditions. Smith et al. demonstrated how, when using this 1x4 testing design, a non-XSL learner known as the *one-exposure learner* could theoretically outperform the cross-situational learner (Smith et al., *under revision*). The one-exposure learner is a learner who is only exposed to a single trial condition or who only remembers the details of a single trial. The one-exposure learner is not using XSL, and his performance on the test depends on the overlap of training foils in the single trial condition and in the test condition. Using a mathematical calculation, Smith et al. (*under revision*) show that the one-exposure learner outperforms Yu and Smith's (2007) participants in the 3x3 and 4x4 conditions. Smith et al. argue that a more accurate XSL test would include all possible foils. Smith et al. replicated Yu and Smith's study using the remediated design, and found that XSL was less robust than previously thought. In particular, adults could learn the 18 words with 6 exposures per word, but only in the 2x2 and 3x3 conditions. Thus, Smith et al. concluded that large sets of words could be learned simultaneously and rapidly but only when using low levels of referential uncertainty. The authors mention that their results underscore the importance of word-learning heuristics in language acquisition because they significantly constrain referential uncertainty. This statement connects back to the idea of an integrative account of word learning where the learner is processing multiple cues (e.g. Hollich et al., 2000).

3.1.3. Smith et al. paradigm

Smith et al. (2010; *under revision)* have run XSL experiments in the laboratory using a paradigm that also allows for controlled degrees of referential uncertainty. The Smith et al. paradigm varies and quantifies referential uncertainty using a 1x*C* condition, where *C* can be 2, 3, 4, 5, etc. (Smith et al., 2010). Each trial has only 1 label but multiple referents. Each word appears with a specific number of distractors that are not targets or distractors for other words. This is different from Yu and Smith's (2007) experiment, where distractors for one word are target referents for another word. The Smith et al. paradigm also uses a repeated testing approach, such as in the HSP, where a test exposure occurs after each training exposure. Figure 3.3 illustrates one training and one testing phase used in Smith et al.'s (2010) experiment.

Smith et al. (2010) ran an XSL experiment where each target word was associated with a total of 15 possible referents, and each training trial had 2, 5, or 8 of the 15 referents shown. The authors varied the mode of presentation so that half of the words were presented consecutively while the other half were interleaved with the other target words. Participants were said to have learned a word if they chose the correct referent in the final test exposure. Overall, Smith et al. found that a significantly higher number of participants were able to learn the words in both the consecutive and interleaved blocks compared to the number of learners that would be expected to learn the words if using the one-exposure learning strategy. However, there are several important results to note. Firstly, accuracy rates decreased as *C*, the level of referential uncertainty, increased. This was true for both the consecutive and interleaved presentation modes. Secondly, there was a significant effect of *C* on the learning time in both presentation modes. An increased *C* resulted in an increased learning time. Thirdly, there was no significant interaction between degree of referential uncertainty and presentation mode, which suggests that adults are as capable of learning words when interleaved with other words as when displayed consecutively. Smith et al., in accordance with Yu and Smith (2007), conclude that XSL in adults is possible even with high levels of referential uncertainty.

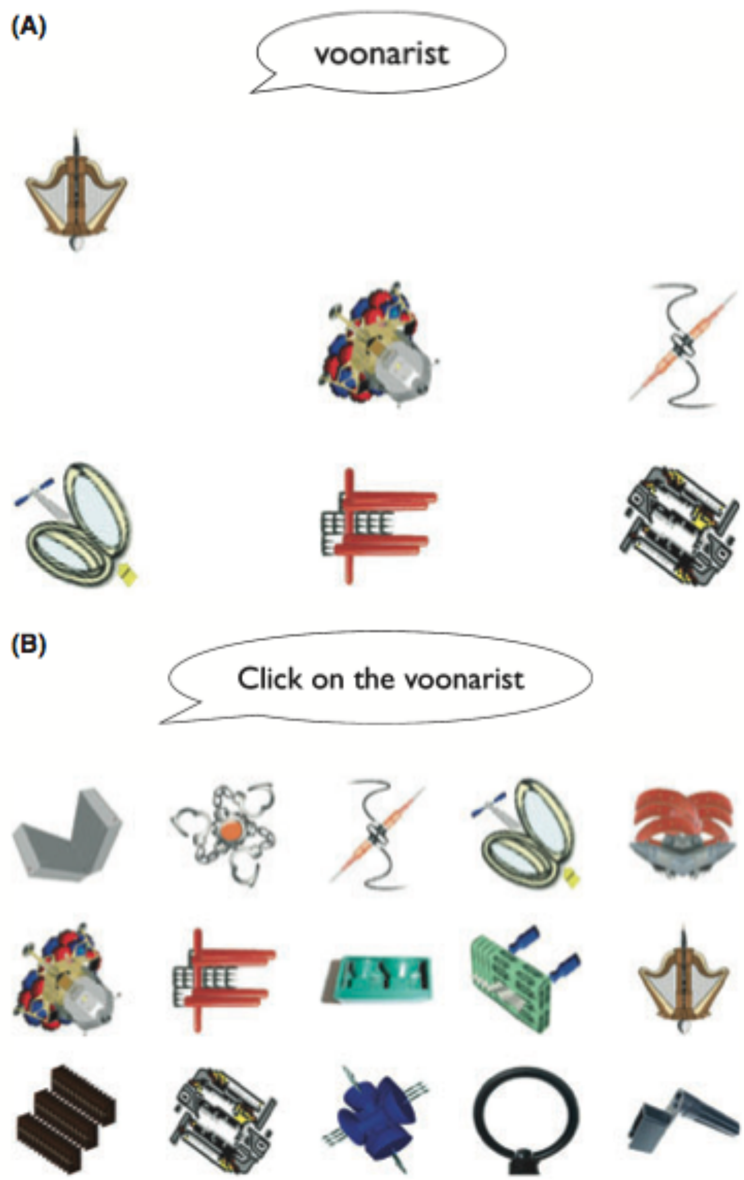


Figure 3.3: Exposures from the training phase (*a*) and the testing phase (*b*) used in Smith et al. (2010). In this example, Smith et al. are using a 1x5 design for the training phase, and the testing phase consists of all 15 possible referents. This image is reproduced from Smith et al. (2010).

In summary, Sections 1.2.1, 1.2.2 and 1.2.3 have described several XSL experiments examining learning both noun and verb learning, and using various paradigms. It is clear that XSL is possible, and even considered robust (Yu & Smith, 2007), in adults, children, and infants.

3.2. BEHIND XSL: ASSOCIATIVE LEARNING MECHANISM

The empirical work on XSL has often been accompanied by theoretical work on the possible mechanisms behind the tracking of co-occurrence frequencies that occurs in XSL. Initially, two main XSL mechanisms were proposed: a simple, Hebbian-like associative model (Yu & Smith, 2007) and a more complicated competitive, hypothesis-testing model (Siskind, 1996; Xu & Tenenbaum, 2007[1]). In the simple associative learning model, co-occurrences of all words and objects are consistently being computed and word-object associations are strengthened with each additional co-occurrence of the word and object (Yu & Smith, 2007). On test trials, the model selects the pair that has most frequently co-occurred.

In the hypothesis-testing model, proposals are created about specific word-object pairs, and altered and stored in memory with each trial. If future trials are consistent with the initial proposal, the proposal is retained. If future trials are not consistent with the proposal, it is eliminated from the lexicon (Siskind, 1996). A simple associative model is consistent with a largely implicit or incidental process of learning where learners are sensitive to co-occurrences without explicitly directing attention to word-object pairs, while a hypothesis-testing model is more compatible with a deliberative, explicit learning strategy (Kachergis, Yu, & Shiffrin, 2010).

Since the initial proposals of two different types of learning mechanisms, more work has been done to give further insight into XSL learning mechanisms. Smith et al. (2010) found that there were two distinct XSL strategies being used: pure XSL and approximate XSL. Pure XSL learners were keeping track of the co-occurrence frequency between a word and all possible referents, while approximate XSL learners were keeping track of a select number of word-referent pairs based on the current and previous contexts. Smith et

---

[1] Tenenbaum and Xu (2000) discuss a Bayesian inference model of word learning, where learners are evaluating word-referent probabilities while taking prior and posterior probability into account.

al. propose a continuum of learning strategies where pure XSL and approximate XSL are varying degrees of an associative learning device.

In summary, several mechanisms behind XSL have been suggested but no conclusive answer has been reached. It has also been suggested that an associative learning mechanism is more important in infancy and helps guide infants to be able to use a more direct and hypothesis-based mechanism for word learning later in life (Smith & Yu, 2008).

3.3. TESTING FOR MULTIPLE CUES USING XSL PARADIGMS

Several XSL studies have begun to incorporate different aspects of natural languages into XSL studies to measure the extent to which various factors aid XSL in increasingly complex conditions. These studies often incorporate several different types of cues, such as statistical and linguistic cues. The following subsection will look at studies that have incorporated linguistic cues and prior knowledge in XSL studies.

3.3.1. Bootstrapping in language acquisition

Bootstrapping is a well-known and discussed phenomenon in language acquisition that has been used to explain children's abilities to acquire language quickly and relatively effortlessly (e.g. Gillette et al., 1999; Gleitman, 1990; Pinker, 1989).

As mentioned in Chapter 2, there are two traditional bootstrapping hypotheses: syntactic bootstrapping and semantic bootstrapping. The theory of syntactic bootstrapping (Lidz, Gleitman, & Gleitman, 2004; Gleitman, 1990) states that the children use knowledge of sentence structure to deduce meanings of words. When observing real-world situations, the child also observes the specific structures in which the words occur, and then uses this knowledge when confronted with novel words. Gleitman (1990) pays a special attention to verb learning and posits that there is an important relationship between verb meaning and syntactic structure. Experiments using Gillette et al.'s HSP (1999) found that syntactic information significantly improved verb learning.

Semantic bootstrapping (Pinker, 1987; 1989) is the reverse of syntactic bootstrapping. The semantic bootstrapping hypothesis states that children acquire a sizable mental lexicon and then use the syntactic-semantic links to determine the syntactic category of the words for which they have concepts. Pinker's theory of semantic bootstrapping is used to support an innate capacity for language. In particular, children have an innate knowledge of syntactic structure, and use their initial knowledge of several words to constrain possible word argument structures and to determine subcategorization frames (Pinker, 1984).

Both the syntactic and semantic theories of bootstrapping support the idea of using multiple cues in order to either facilitate word learning or syntax learning and can thus be considered the beginnings of an integrative theory of word learning. However, it is important to mention that these theories invoke the idea of *innate* language-specific knowledge that helps them establish syntax-semantics mappings (Bloom, 1994; Gleitman, 1990; Pinker, 1989). The effects of bootstrapping can be examined without relying on the innateness argument. The following subsections discuss several XSL studies that have integrated multiple cues into the experimental or computational designs.

3.3.2. Integrating bootstrapping into XSL experiments

Several recent XSL experiments have focused on learning nouns that are embedded in sentences in order to examine the effect of syntactic bootstrapping (Alisha & Fazly, 2010; Köhne and Crocker 2010; Yu, 2006) and prior knowledge (Yurovsky et al., 2009) on word learning during an XSL task. These studies either directly or indirectly demonstrate how multiple cues, if available to the learner, are used to improve word learning.

Köhne and Crocker (2010) studied the role of sentence processing mechanisms during XSL of nouns by tracking the eye movements of adult participants. In the experiment, participants learned nonsense nouns that were embedded in whole sentences. The authors were interested in seeing whether knowledge of the syntactic frame of a previously learned verb would aid participants in novel noun learning by restricting the number of possible referents. This is a sentence processing mechanism that occurs when people use their native languages. In Köhne and Crocker's experiment, XSL was referred to as "visual context" and syntactic bootstrapping as "linguistic context". Participants were taught 6 restrictive, transitive verbs in an alien language, and then were exposed to a set of XSL training trials using whole sentences consisting of two determiners, one character (subject), one verb, and one object. The verbs learned prior to the XSL experiment were either clothing verbs (*melimema*, 'iron') or food verbs (*bermamema*, 'eat'), while the nouns to be learned using XSL were either human characters (*badut*, 'clown'), food items (*sonis*, 'sausage'), or clothing items (*oblung*, 't-shirt'). Each training trial depicted the target agent (subject) and target patient (object), as well as an additional distractor agent and distractor patient. Additionally, there was always one food item and one clothing item. A strict SVO order was used, although participants were not informed of this. After the training trials, participants were given a forced-choice vocabulary test that tested for knowledge of the nouns. Köhne and Crocker found that participants looked more consistently at characters than at subjects when they heard the first noun of the sentence and at objects than at characters when they heard the second noun. Additionally, they had more reliable inspection of the target object than of the distractor object. Köhne and Crocker concluded that participants were able to use native-like sentence-processing mechanisms to guide their learning. Although Köhne and Crocker's study suggests that two types of cues, statistical and linguistic, can be used in XSL experiments, it contains two possible confounding factors. Firstly, their study used common nouns and verbs for which all participants would have had pre-established concepts, which renders their study more applicable to foreign language learning. Secondly, participants were *taught* the set of verbs before being exposed to XSL trials for noun learning. Participants were therefore going into the task with a large amount of prior knowledge that they did not acquire through XSL.

Alishahi & Pyyukkönen (2011) used a probabilistic computational model of word learning to examine the effect of syntactic bootstrapping. The model was presented with a scene-utterance pair along with additional syntactic information, such as the head words for the main predicate and its arguments, the word order, the number of arguments, and the conceptual characteristics of the event (e.g. *cause*, *change*). In the model, a base knowledge of lexical categories was assumed. The model functioned by calculating probabilities based on cross-situational observation and syntactic observation. Alishahi & Pyyukkönen found the syntactic information helped verb identification but not noun identification. They also found that the onset of syntactic bootstrapping was delayed, and they interpreted this as syntactic bootstrapping occurring later in life, once a child has a stable knowledge of syntax in place. Alishahi and Pyyukkönen's study offers support for syntactic bootstrapping in word learning. However, the input presented to the model lacked referential uncertainty and their results may therefore not be reflective of results in XSL experiments.

There are some XSL experiments that also refer to bootstrapping through prior knowledge (Klein, Yu, & Shiffrin, 2008; Yurovsky et al., 2010). It has been empirically demonstrated that children use their own prior knowledge to help them fast-map the names of new nouns. Prior knowledge is particularly important when children employ a mutual exclusivity bias (Au, 1990). It is important to note that this use of "bootstrapping" does not refer to innate knowledge, but rather knowledge that has been previously gathered through empirical means. A recent study by Yurovsky et al. (2010) examined the role of partial, sub-threshold, and often incomplete, knowledge in learning. Yurovsky et al. (201) ran an experiment where participants were exposed to two training and testing sets. The *incorrect* answers from this test were used as stimuli in the second set. Yurovsky et al. (2010) reasoned that partial knowledge should help constrain the possibilities in the XSL experiment, reduce ambiguity, and thus boost learning. It was found that participants in the experimental group learned twice as many words as the participants in the control group, who had all new stimuli in the second set. Yurovsky et al. (2010) concluded that sub-threshold partial knowledge drove the learning of both old and new words.

3.4. SUMMARY OF BACKGROUND INFORMATION

Chapters 2 and 3 have described several different accounts of word learning, and have emphasized the power of correlating cues specifically in lexical development and, more generally, in language acquisition. Chapter 3 focused on two cues, co-occurrence frequencies and syntactic-semantic links, examining them as separate cues and as mutually reinforcing cues. Cross-situational learners are sensitive to co-occurrence frequencies between a word and its referent and can extract this information over multiple exposures (Siskind, 1996; Smith et al., 2010; Yu & Smith, 2007), and language learners can use syntactic-semantic links to bootstrap lexical (Gleitman, 1990) or syntactic development (Pinker, 1987). The ability to use syntax-semantics links in an XSL experiment has been hinted at through an empirical study (Köhne & Crocker, 2010) and suggested by a computational model (Alishahi & Pyyukkönen, 2011). These studies combine two accounts of word learning: XSL (e.g. Siskind, 1996) and syntactic bootstrapping (e.g. Gleitman, 1990) and conclude that syntactic bootstrapping can aid XSL. The results are compatible with the idea that multiple cues reinforce each other and enable faster lexical development (e.g. Hollich et al., 2000). In chapter 4, an original research experiment that examines the combined effect of XSL and syntactic bootstrapping will be presented. The results will be used as additional support for an integrative account of word learning.

CHAPTER 4: ORIGINAL XSL RESEARCH EXPERIMENT

Previous XSL studies using the Yu and Smith or Smith et al. paradigm have often focused only on nouns (e.g. Klein et al., 2008; Yu & Smith, 2007; Smith & Yu, 2008; Smith, Smith, & Blythe, 2010; Yurovsky et al., 2010) or on verbs (Scott, 2010), without combining the two grammatical categories and without incorporating additional cues into the experiments. However, Köhne and Crocker (2010) provided an experimental study and Alishahi and Pyyukkönen (2011) a computational study that suggested that the syntactic framework of a sentence could aid XSL through a bootstrapping process. The present experiment aimed to provide an XSL experiment of *sentences* that clearly tested for the combined effects of XSL (Siskind, 1996) and syntactic bootstrapping (Gleitman, 1990) on word learning. The experiment incorporated statistical cues (co-occurrence frequencies) and syntactic cues (syntactic-semantic links that allow for bootstrapping).

In the current experiment, adults were trained and tested on a total of 15 nouns and 15 verbs, grouped into three sets of 5 nouns and 5 verbs. Each set represented a different condition. One set used an unstaggered learning method, where all nouns and verbs are presented and learned in one *training* block. The other two sets involved a staggered training method, where the first training block, known as the *pretraining* block, involved cross-situationally learning only either nouns or verbs, and the second block, known as the *training* block, involved learning both nouns and verbs together. The staggered training method allowed for the opportunity to examine the use of prior knowledge acquired from the *pretraining* block and transferred into the *training* block. The staggered training also reduced the dependence on co-occurrence frequencies in the training block and allowed for the use of syntactic knowledge.

Several questions were addressed in this study:

1. Can adults simultaneously cross-situationally learn nouns and verbs when both are embedded in two-word sentences?
    a. Are they learning one of the grammatical categories better than the other?

2. Does staggered learning allow for syntactic bootstrapping and thereby the facilitation of word learning?
    a. If staggered learning does improve scores, which initial XSL learning provides more support for later XSL learning: nouns or verbs?

Based on prior discussions regarding the robustness of XSL (e.g. Yu & Smith, 2007; Smith et al., *under revision*) it was hypothesized that simultaneous XSL of nouns and verbs in a two-word sentence would indeed be possible. In regards to the bootstrapping hypotheses, it was hypothesized that, in the staggered training sets, participants would be able to integrate prior knowledge from the *pretraining* block and syntactic knowledge to constrain the possible word-referent pairs in the *training* block and to improve accuracy scores and response times on the vocabulary tests. Finally, because previous studies have suggested that verbs are more difficult than nouns to be learned (Gillette et al., 1999; Piccin & Waxman, 2007), it was hypothesized that learning verbs first would reduce more of the workload in the XSL than learning nouns, and would lead to higher accuracy and faster reaction times.

4.1. METHOD

4.1.1. Subjects

This study recruited 51 students from the University of Edinburgh (25 females). All participants had normal or corrected-to-normal vision. Participants were recruited from the University of Edinburgh Career Service database or from emails directed to the school of Philosophy, Psychology, and Language Sciences at the University of Edinburgh. Each participant was paid £3 for his or her participation. None of the participants had previously taken part in an XSL experiment. As the study used nonwords, it did not discriminate between native English or non-native English speakers, nor between monolingual and bilingual speakers. There were two versions of the experiment; 28 participants were given version A (14 females) and 23 were given version B (11 females). Data from 4 participants in the first version A and 5 participants in the second version were discarded due to technical difficulties during the experiment, so that the final number of participants whose results were analyzed for both versions was 42 (22 females); 24 of these completed version A and 18 completed version B.

4.1.2. Stimuli

This study used novel and distinct nouns and verbs spoken and written in an alien (nonsense) language. Initially, there were a total of 18 novel objects and 18 actions created for the experiment, with 6 objects and 6 actions per condition. However, due to technical difficulties,[2] the number of objects and actions was reduced to 15, where each condition had 5 verbs and 5 nouns, with no overlap between conditions. This yielded 25 distinct animations per condition, and a total of 75 distinct animations overall for each version. The objects were a subset of the objects used in Smith et al. (2010). They were made from splicing real (e.g. bicycle, lamp) and artificial objects and rearranging them to form novel objects. All verbs were intransitive. For a full list of the nouns and the objects, see Appendix I.

The nonsense names describing the 15 novel nouns and 15 verbs were created using the English Lexicon Project Website (Balota, Yap, Cortese, Hutchinson, Kessler, Loftis, et al., 2007). All nonsense words (both nouns and verbs) followed English phonotactic rules and were 1, 2, or 3 syllables in length and they varied in syllable onset (vowel, single consonant, or consonant cluster). To eliminate the possibility of using known English grammar to help learn the words or to find patterns within the pseudowords, typical verbal inflections in English (e.g. *–ing*, *-ed*) were not used and the endings of the nouns were matched with the endings of the verbs (i.e. whenever one noun ended with "-*y*", there was also one verb that ended with "-*y*"). Spoken forms of the nouns and verbs were recorded with the voice of male, native English speaker. The words were recorded in sentences and included typical sentential prosody.

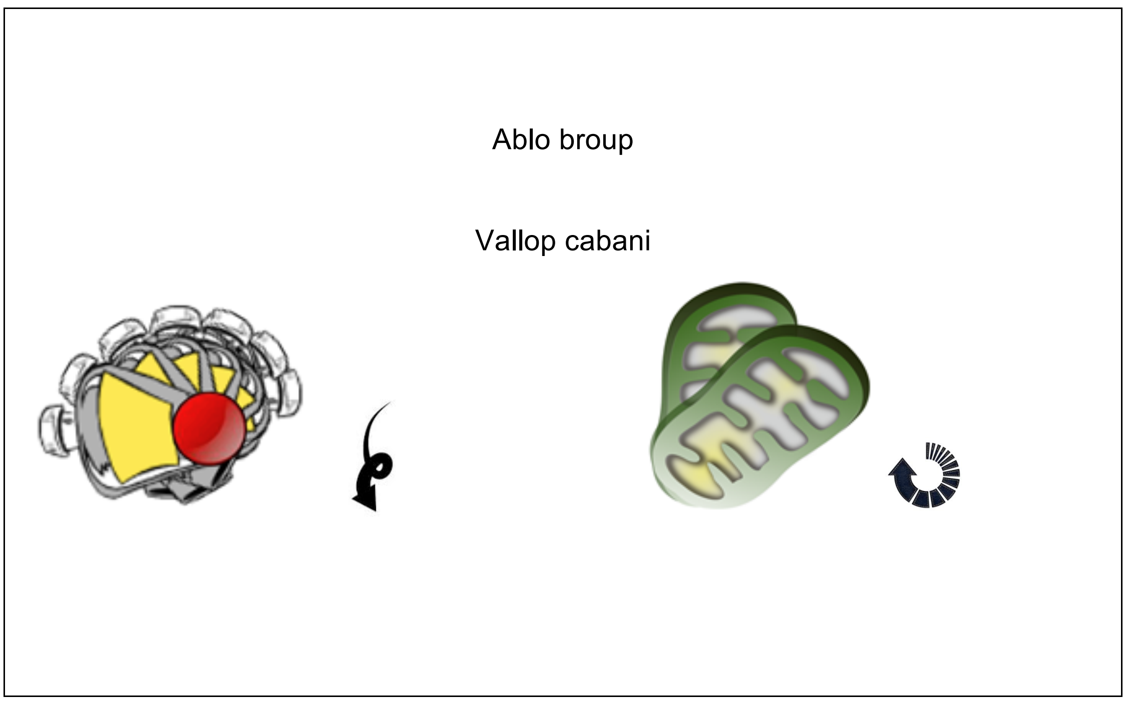
4.1.3. Design

The nouns and verbs were combined into two-word sentences. A free word order was used for the sentences, and the position of nouns and verbs was pseudo-randomized

[2] The software used for the experiment, E-Prime 2.0 was unable to play 6 videos simultaneously for the testing phases. For this reason and to avoid training on a noun-object and verb-action pair without being tested on it, the total number of videos was reduced to 5.

so that half of the sentences were SV (noun-verb) and half were VS (verb-noun). The word orders were interleaved randomly in both the *pretraining* and *training* blocks. Each *training* trial consisted of two sentences, presented auditorily and visually, and two animations. The level of referential uncertainty was therefore 2x2 for all participants in each condition, although a 2x2 design refers to 2 noun-object pairs and 2 verb-action pairs. The set up is therefore similar to Yu and Smith's (2007) 4x4 design, but the referential uncertainty is not as high as in Yu and Smith's experiment due to the grouping of words into sentences. An example of the set up for a *training* and testing block is illustrated in Figure 4.1. For each *training* block, the correct word-referent pair occurred 6 times per block. For a list of all word-referent probabilities, see Appendix II.

a)



b)

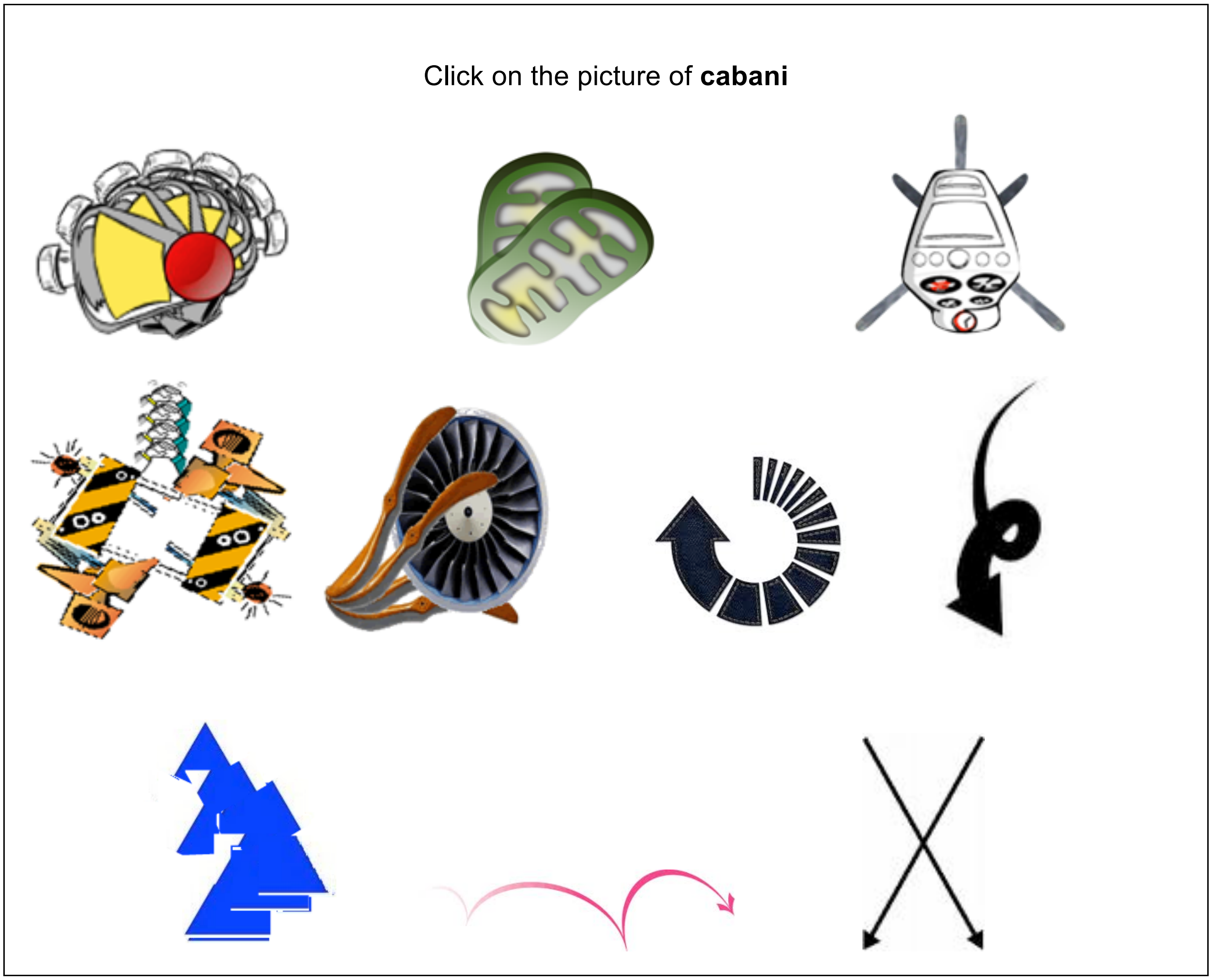


Figure 4.1: An example of the *training* block (*a*) and testing block (*b*). The arrows indicate actions. In the experiment, the arrows were replaced with an animation of a blue triangle performing the action. In the *training* block, either sentence could refer to either object, and within each sentence, either word could refer to the noun or the verb. In the testing block, the blue picture on the bottom left is a snapshot of the blue triangle performing the animation "morph".

All participants were tested on all three conditions: the unstaggered condition (henceforth Unstag), the staggered condition with pretraining on nouns first (henceforth StagNoun), and the staggered condition with pretraining on verbs first (henceforth StagVerb). The order of the sets was counterbalanced so that each version had 6 counterbalanced versions, and the counterbalanced versions were equally distributed among the participants. The experimental design is summarized in Table 4.1.

Condition	Block	Number of words learned	Type of words learned	Total Number of occurrences	Number of Trials	Time per trial (s)	Total time (s)
Unstaggered (Unstag)	Train	10	Nouns + verbs	6	15	10	150
Staggered-with-nouns-first (StagNoun)	Pretrain	5	Nouns			5	75
	Train	5	Nouns + verbs			10	150
Staggered-with-verbs-first (StagVerb)	Pretrain	5	Verbs			5	75
	Train	5	Nouns + verbs			10	150

Table 4.1: Experimental conditions. All conditions used a 2x2 level of referential uncertainty.

4.2. PROCEDURE

Participants were tested individually using a Toshiba 15" laptop. Headphones were used to listen to the auditory stimuli. Participants were told that they would be shown a series of computer slides that consisted of two, two-word sentences and two animations and that their task was to learn the names for the objects and actions. They were not explicitly told to use an XSL approach. Each participant was exposed to the three different conditions: Unstag, StagNoun, and StagVerb. The appearance of the stimuli was pseudorandomized in order to avoid two of the same nouns or verbs appearing together within one trial. Both auditory and visual presentations of the nouns and verbs were used.

The set with the Unstag condition consisted of only one block, the *training* block, in which all stimuli appeared at the same time. Participants were instructed to learn objects and actions simultaneously. There were a total of 15 trials, and each word-target pair appeared 6 times. 2 sentences consisting of a noun and a verb and 2 animations depicting an object and an action appeared on screen. There was ambiguity within each trial but regularity between trials. In this condition, there was little predetermined opportunity for bootstrapping because participants had no prior knowledge of the names of a set of words. Their initial and primary learning method would have been XSL.

The StagNoun condition consisted of two blocks. In the *pretraining* block, participants were initially exposed to all of the nouns only, and had the opportunity learn the noun-object pairs before being exposed to both nouns and verbs. The *pretraining* block replicated the Yu and Smith (2007) noun XSL learning experiment in the 2x2 condition, and had 15 trials, with each noun-object pair appearing 6 times. The objects shown to the participants were still images; no movement was involved. This ensured limited verb interference during noun learning. In the *training* block, participants were exposed to both nouns and verbs. The *training* block ran identically to Condition 1, with each word-target pair appearing 6 times over 15 trials. The only difference was in the initial knowledge the participants might have accumulated from the *pretraining* block that could help render the *training* block easier through exploiting prior knowledge and syntactic-semantic links. This is one of the two conditions that tested for the combined effect of syntactic bootstrapping (through staggered learning) and XSL.

The StagVerb condition ran identically to the StagNoun condition, except that participants learned the 5 novel *verbs* in the *pretraining* block. In order to present the actions in the most straightforward way possible, all animations that depicted only a verb used a neutral blue triangle to perform the action. Participants were explicitly told to focus on the *action* that the object was performing, rather than the object itself. A neutral object that performed all of the actions was used to ensure limited noun interference during verb learning. This is the second condition that tested for the combined effect of bootstrapping and XSL

Theoretically all of the conditions were designed to include both statistical and syntactic cues, because the words were organized into sentences for all conditions. However, it was thought that the Unstag condition provided a mode of presentation that was unfavourable for the use of syntactic bootstrapping because the participants had no initial semantic knowledge. Given the high degree of referential uncertainty, participants in the Unstag condition were obligated to pay attention primarily to the statistical cues.

Forced-choice vocabulary tests were administered to all participants after the *training* blocks in all three sets and additionally after the *pretraining* blocks in the two staggered conditions. During the test after each *training* block, participants heard and were shown one word, and were asked to select its referents out of a possible 10 referents (5 objects and 5 actions). The objects were presented as still images and the actions were presented as animations performed by the neutral blue triangle. The *pretraining* test design was identical to the *training* test design except there were only nouns or verbs and therefore a total of 5 possible referents. Participants had a maximum of 30 seconds to select the right answer before the next selection appeared.

4.3. RESULTS

An initial t-test revealed that the accuracy for *all* three learning types was significantly above chance (Unstag: $t(41) = 6.34$, $p < .01$; StagNoun: $t(41) = 27.00$, $p < .01$; StagVerb: $t(41) = 17.678$, $p < .01$). Figure 4.2 illustrates the mean accuracy proportions for all learning types.

A two-way repeated measures ANOVA compared the effects of learning type, word type (within-subject factors) and version type (between-subject factor). According to Mauchly's test of sphericity, the assumption of sphericity was violated when testing for learning type in the two-way ANOVA ($\varepsilon = .68$ for the main effect of learning type). Therefore, the Greenhouse-Geisser corrected degrees of freedom were reported. Significant main effects were found for learning ($F(1.513, 60.534) = 74.689$, $p < .001$, $\eta^2 = .651$) and for word type ($F(1, 40) = 10.256$, $p < .01$, $\eta^2 = .204$), with nouns being identified significantly more than verbs. There was no main effect for version type, ($F(1.813, 60.534) = 1.465$, $p > .05$); consequently, in many of the subsequent tests, the two groups were collapsed into and analyzed as one group of 42 participants. Importantly, there was also a significant interaction effect between learning type and word type ($F(2, 80) = 11.189$, $p < .001$, $\eta^2 = .219$).

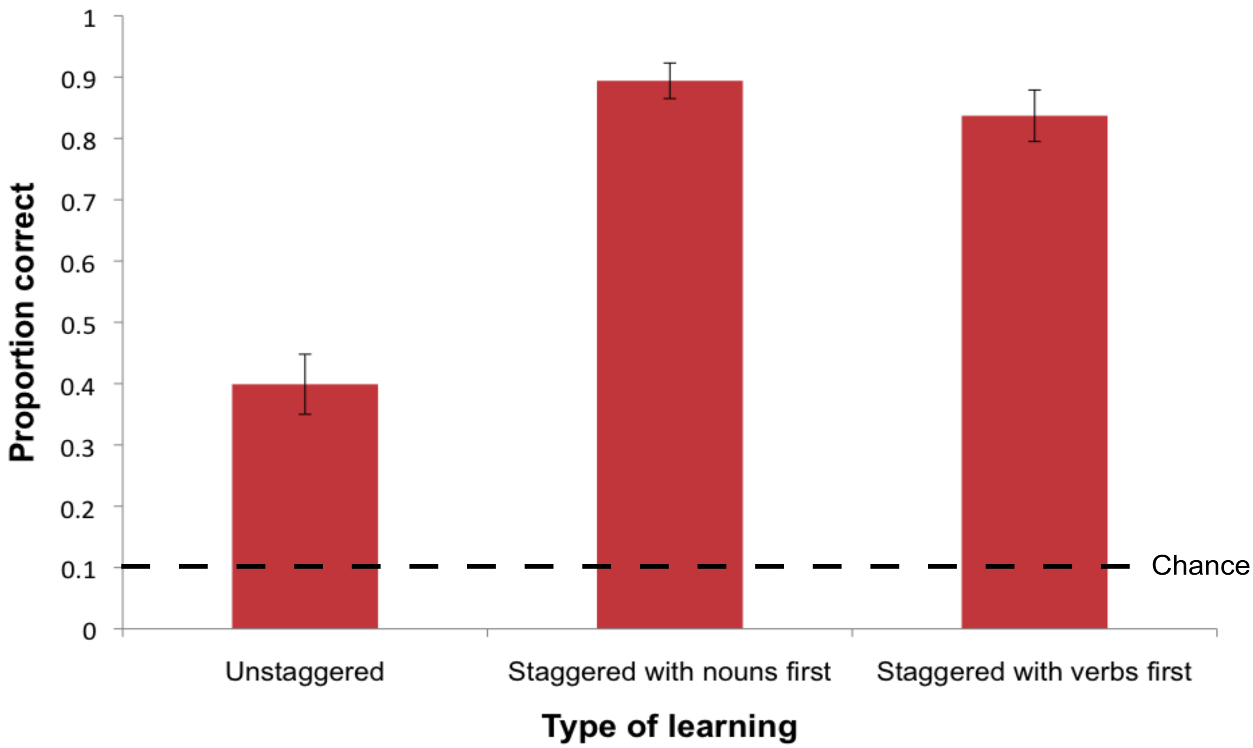


Figure 4.2: Mean performance (total of 10 words per test) on testing conditions based on learning conditions. Error bars give the ANOVA standard error calculations. Dashed line shows chance levels.

Post hoc comparisons using Tukey's HSD test indicated that the mean performance in the StagNoun (*Mean difference (MD)* = .479, $p < .001$) and the StagVerb (*MD* = .429, $p < .001$) was significantly higher than in the Unstag condition. No significant difference was found between the StagNoun and StagVerb condition (*MD* = .05, $p > .05$).

In order to further compare the StagNoun and StagVerb conditions, a two-way repeated measures ANOVA was run to compare reaction times of the different learning conditions (within-subject variables) and version type (between-subject variables). Once again, a significant main effect was found for learning ($F(1.059, 42.356) = 6.308$, $p < .05$, $\eta^2 = .136$). Once again, the Greenhouse-Geisser corrected values were used (ε .111). Post hoc comparisons using Tukey's HSD indicated that reaction times were significantly lower for the StagNoun (*MD* = -2808.815 ms, $p < .05$) and StagVerb (*MD* = -3074.712, $p < .01$) conditions compared to the Unstag condition, but the StagVerb condition was not significantly lower for the StagVerb condition than for the StagNoun condition (*MD* = -265.890, $p > .05$).

A separate ANOVA comparing the *differences* between accuracy for nouns and accuracy for verbs for each learning condition found a significant main effect of the difference in accuracy for learning type ($F(2, 80) = 11.189$, $p < 0.05$). This confirmed the results of the initial two-way ANOVA that revealed a significant main interaction effect. Figure 4.3 illustrates the differences in means for each condition. Post hoc tests using Tukey's HSD test indicated that the mean difference in noun and verb accuracy was significantly higher in the Unstag condition than in the StagNoun (*MD* = -.143, $p < .05$) and in the StagVerb (*MD* = -.238, $p < .001$) conditions. However, the mean difference in noun and verb accuracy between the StagNoun and StagVerb condition was not significant (*MD* = .095, $p > .05$).

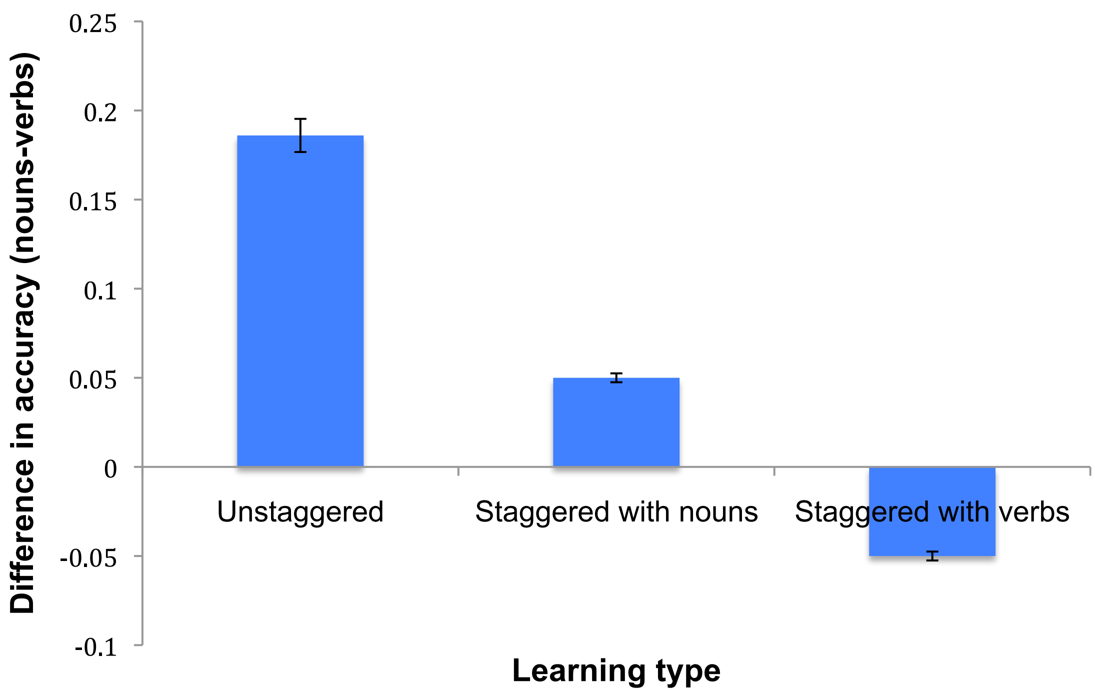


Figure 4.3: Difference in the mean accuracy for nouns and the mean accuracy for verbs in each condition. Error bars give the 5% error rate.

Several post hoc tests using Tukey's HSD test were performed to examine the accuracy independently for nouns and verbs within one condition and between conditions. Participants performed significantly better on nouns than verbs in the Unstag condition (MD = .190, $p < .05$), but there were no significant differences between noun and verb performance in the StagNoun (MD = .048, $p > .05$) or in the StagVerb (MD = -.048, $p > 0.05$) conditions. Performance on *nouns* was significantly higher in the StagNoun condition (MD = 0.410, $p < .001$) and StagVerb condition (MD = .324, $p < .001$) than in the Unstag condition. Performance on *verbs* was significantly higher in the StagNoun condition (MD = .552, $p < .001$) and in the StagVerb condition (MD = .562, $p < .001$) than in the Unstag condition. No significant differences between StagNoun and StagVerb conditions were found for noun accuracy (MD = .085, $p > .05$) or for verb accuracy (MD = -.01, $p > .05$).

Finally, a paired t-test on the two pretraining tests found no significant difference in performance between the StagNoun pretest and the StagVerb pretest ($t(41) = 1.120$, $p > .05$). It should be noted that the t-test was comparing performance on *nouns* in the StagNoun condition with performance on *verbs* in the StagVerb condition.

4.4. DISCUSSION

The results of the experiment indicate that adults *can* simultaneously cross-situationally learn nouns and verbs that are embedded in two-word sentences when co-occurrence frequencies are the only cue, but also that they perform much better when given the opportunity for syntactic bootstrapping. Although accuracy in the Unstag condition was significantly higher than chance, adults were still only correctly identifying less than half of the words. There was a dramatic increase in accuracy when staggered learning was introduced in the experiment. Mean accuracy scores increased from under 5/10 to nearly 9/10 for the StagNoun condition and 8.5/10 for the StagVerb condition. 15, 5-second exposures of just nouns or just verbs in the *pretraining* block significantly increased the accuracy results on the *training* block for both nouns and verbs. This provided direct evidence of effect of prior knowledge and syntactic bootstrapping on word learning.

The effects of bootstrapping through prior knowledge were seen when comparing the performance on nouns in the Unstag and StagNoun conditions and the accuracy of verbs in the Unstag and StagVerb conditions. In both cases, performance was significantly increased in the both staggered conditions. Participants in the staggered conditions received an additional training set, the *pretraining*, on the set of words being tested. That is to say, they saw either the nouns or the verbs an additional 6 times in the *pretraining* blocks, and therefore saw the nouns or verbs, depending on the condition, 12 times compared to only 6 times in the Unstag condition. These findings are consistent with Yurovsky et al.'s (2010) results which indicated that partial knowledge in an XSL experiment significantly helped acquire future knowledge.

Syntactic bootstrapping (as defined by Gleitman, 1990) can be seen when comparing noun accuracy in the Unstag condition with the StagVerb condition, and verb accuracy in the Unstag condition with the StagNoun condition. Once again, the accuracy was significantly higher in the staggered conditions. However, in these cases, the participants received the same amount of training as in the Unstag condition, with respect to the individual grammatical category. Nouns appeared only 6 times in the StagVerb condition and verbs appeared only 6 times in the StagNoun condition, which is the same number of times nouns and verbs appeared in the Unstag condition. These experimental results suggest that knowledge of the word-referent pair for a word in one grammatical category restricted the possible word-referent pairs for the other grammatical category. These results are compatible with Köhne and Crocker's (2010) that examined sentence-processing mechanisms during an XSL task, although, unlike Köhne and Crocker's experiment, the results do not indicate whether participants were using the same sentence processing mechanisms as they use in their first languages. Furthermore, the results are compatible with the idea of syntactic bootstrapping through the use of syntax-semantics links to restrict the possible word-referent pairs in a sentence (Gleitman, 1990).

Conclusively, then, bootstrapping had a significant effect on XSL in two ways: through syntactic-semantic links and through prior knowledge. Without the ability to bootstrap knowledge, participants were scoring under 50% on the recognition tests. With the ability to bootstrap, this score was raised to over 80%.

These results suggest that co-occurrence frequency, syntactic framework, and prior knowledge were all helping the participants learn the novel words, and that syntactic bootstrapping and XSL was more effective than XSL alone. The results corroborate the emergentist coalition model of word learning, which states that word learning occurs through the extraction and interpretation of multiple cues (Hirsh-Pasek et al., 2000). Other researchers have provided additional support for the joint acquisition of different types of cues during. For example, Maurits, Perfors, & Navarro (2009) a computational model that acquired word order and word reference jointly They found that the joint acquisition rendered both language problems easier because they were mutually constraining each other.

More generally, the conclusions drawn from this study, that a diverse set of cues, including statistical, linguistic, and social cues, facilitate word learning, can be applied to *all aspects of language acquisition* (Aslin & Newport, 2008; Cunillera, et al., 2010a; Romberg & Saffran, 2010; Sahni, Seidenberg, & Saffran, 2010). There are several ways in which multiple cues facilitate language acquisition. Multiple cues can direct attention; mark seemingly less salient structures, such as nonadjacent dependencies (Newport & Aslin, 2004); and lead to bootstrapping, where one cue enables recognition of a second cue that eventually becomes more important than the first cue (Romberg & Saffran, 2010; Sahni et al., 2010).

One interesting and unexpected result was that participants performed significantly better on identifying nouns than verbs in the Unstag condition, but performed equally well on the two grammatical categories in both staggered conditions. This latter result is unexpected because there was increased training on verbs but not on nouns in the StagVerb condition, and on nouns but not on verbs in the StagNoun condition, yet there was no difference in performance between the two grammatical categories for either condition. A possible explanation for the discrepancy between performance on nouns and verbs in the Unstag condition could be the difference in imageability between nouns and verbs. In the HSP experiments, which involved XSL of nouns and verbs, three times as many nouns as verbs were identified (45% of nouns compared with 15% of verbs) (Gillette et al., 1999). It was speculated that the discrepancy found in the HSP results was due to the facts that verbs are less imageable than nouns and that learning of verbs requires additional linguistic information, such as its syntactic frame (Gillette et al., 1999; Piccin & Waxman, 2007). Although a useful explanation for Gillette et al's study, it is likely that this was not the case for the current study, because no difference in accuracy was found between the nouns in the StagNoun pretest and the verbs in the StagVerb pretest. Another explanation for the discrepancy between noun and verb performance in the Unstag condition but the lack of discrepancy in the StagNoun and StagVerb conditions is that there was interference from nouns during verb learning, but not from verbs during noun learning. In the *training* trials, the two grammatical categories were combined to yield 25 distinct animations per condition. Even though the verbs chosen were all distinct, intransitive verbs, such as "moving in a circle", or "growing, then shrinking", participants would have had to separate the movement itself from the object that is moving. It is easier to ignore an action and focus on an object than it is to ignore an object and focus on an action, because the object usually appears as the more salient feature of the animation. The nouns and verbs both appeared 6 times in the *training* sessions. Each time the noun appeared it was clearly visible. On the other hand, each time the verb appeared it was paired with a different salient object. In the StagVerb condition, participants were exposed to a *pretraining* block of just verb learning. The use of a neutral object and the additional 6 exposures to the verbs ensured adequate verb learning. However, in the Unstag and StagNoun conditions, verbs appeared only 6 times and each time with a different action. The learning of verbs thus required the extraction of the relevant and similar feature, the action, over a series of different animations, and may have required more attention and work than was required for the learning of nouns. Despite this being the case for both the Unstag and StagNoun conditions, verb performance only suffered in the Unstag condition. This is possibly because, in the StagNoun condition, participants had previously learned the nouns and therefore may have freed up some working memory to devote to verb learning. In the Unstag condition, there was no prior knowledge and no decrease in work load, and participants may have struggled much more at verb learning, thus allowing noun learning took precedence based on facility. Therefore, it may be more difficult to learn the names for actions when they are being performed by several different objects than when they are being performed by the same object.

One flaw in the experiment is the use of 5 nouns and 5 verbs in each condition rather than the 6 verbs and 6 nouns that were originally planned. Each word-target pair appeared 6 times, which meant that there were a total of 30 pairs: 25 unique pairs and 5 pairs that were seen twice. The use of 5 repeated pairs was unavoidable due to the need to pseudorandomize the appearance of the pairs. Although significant results were still seen, the appearance of some pairs twice could have altered the results by creating stronger links between certain spurious word-referent pairs, especially when the difficulty of learning verbs when paired with different objects is taken into consideration.

One error already mentioned is that all conditions, including the Unstag condition, used 2 sentences and 2 animations, rather than individual 2 noun-object pictures and 2 verb-action animations. The experiment was designed this way to better reflect the structure of natural languages and the type of stimuli child learners are exposed to during word learning as well as to test for *sentential* XSL. Theoretically, all conditions had syntactic cues that were available to the learner, because the stimuli in all conditions were arranged into sentences. However, according to the syntactic and semantic bootstrapping hypotheses (Gleitman, 1990; Pinker, 1989) some initial knowledge is needed in order for bootstrapping to occur. The staggered training method was therefore more conducive than the Unstag method to the use of syntactic because participants had prior knowledge of some words. In the Unstag method, participants began with no initial knowledge of any of the words. Given the high degree of referential uncertainty and the limited time and number of trials in the Unstag condition, syntactic bootstrapping was likely ineffective in this condition. Nevertheless, to create a more precisely controlled experiment that measures a condition with only co-occurrence frequencies as a cue and a condition with additional cues, the current study could be repeated with the addition of another control condition where the words are not visually and auditorily presented as sentences, but rather as individual words without any sentential prosodic cues. This condition could be compared to the current Unstag condition to see the effect of rudimentary syntax. If participants performed worse in this condition than in the Unstag condition in the current study, there would be evidence for prosodic cues and syntactic-semantic links even in the Unstag condition.

A related point to note is that the current experiment used a 2x2 design that was actually more similar to Yu and Smith's (2007) and Smith et al.'s (*under revision*) 4x4 design, because participants were tracking 2 nouns and 2 verbs. In the Unstag condition, mean performance was approximately 40% accuracy. In Yu and Smith's study, mean performance accuracy was 55% accuracy in the 4x4 condition. In Smith et al.'s study, which replicated the Yu and Smith experiment but used a 1x18 rather than a 1x4 testing design, mean performance accuracy in the 4x4 condition was approximately 33% accuracy. Performance in the current study is in between performance in Yu and Smith's and Smith et al.'s studies. This is possibly due to the fact that referential uncertainty was not as high as in Yu and Smith's (2007) and Smith et al.'s (*under revision)* 4x4 conditions because of the grouping of each noun and verb into two-word sentences. If this is true, then syntactic structure may have facilitated word learning even in the Unstag condition. However, conclusive results cannot be drawn due to the differences in the number and type of stimuli used in this experiment compared to Yu and Smith and Smith et al.'s experiments.

There are several future directions that can be taken based on this experiment and topic. One future direction that the current study could take is to recreate sentential XSL studies but to incorporate additional linguistic factors, such as a strict word order. In this case, the degree to which additional linguistic cues aid XSL can be measured.

Another direction to take XSL studies concerns the argument for a domain-general account of language acquisition. XSL is a type of SL, and SL is considered to be a domain-general ability as opposed to a language-specific ability (Saffran & Thiesson, 2007). The idea of domain-generality entails the use of the particular mechanism in other domains, such as purely visual or purely auditory, or tactile (Conway & Christiansen, 2005). There is a growing body of work that supports a domain-general account of language. In particular, two experimental results in SL research argue for a domain-general account of language acquisition mechanisms: the use of the same mechanisms in nonhuman animals, and the use of the same mechanisms for nonlinguistic stimuli in humans (Saffran & Thiesson, 2007). Although XSL uses a relatively basic mechanism of tracking frequencies, there have been no direct tests of XSL ability in nonlinguistic domains in humans or in nonhuman animals. In order to join the growing body of work supporting a domain-general account of language, future studies need to examine the possibility of XSL, using an established paradigm, in nonlinguistic domains and in nonhuman animals.

4.5. CONCLUSION

The current study aimed at measuring and comparing the combined effect of statistical and syntactic cues during word learning. Recent studies have shown that XSL is robust in infants and adults (Smith et al., 2010; Smith & Yu, 2008). Previous studies (e.g. Köhne & Crocker, 2010; Yurovsky et al., 2010) have offered support for the idea that the simultaneous processing of multiple cues facilitates word learning, but have usually taught participants a subset of the words, which may have altered the results because natural learning was not occurring.

This study conducted an XSL experiment that incorporated syntactic bootstrapping. In the study, 42 adults were exposed to two-word sentences where both words have referents; the noun referents were novel objects that participants had previously encountered, and the verbs were distinct intransitive verbs. Participants learned 3 sets of 10 words (5 nouns and 5 verbs) that were presented in a 2x2 design, with 2 sentences and 2 animations. Participants were exposed to three different conditions: unstaggered learning where they simultaneously learned nouns and verbs (Unstag); staggered learning where they learned the 5 nouns in a *pretraining* set and were later presented with nouns and verbs together (StagNoun); and staggered learning where they learned the 5 verbs in a *pretraining* set and were later presented with nouns and verbs together (StagVerb). Three important results were found. Firstly, in all three conditions, performance in the unstaggered condition was significantly worse than in either staggered condition. Participants identified under 50% of the words in the unstaggered condition but over 80% of the words in either staggered condition. The results were interpreted as evidence for bootstrapping through the use of syntax-semantics links and prior knowledge. Secondly, there were no significant differences in accuracy or reaction times between the StagNoun and StagVerb conditions, which indicated that staggered learning worked equally well regardless of what grammatical category was isolated and learned first. Lastly, nouns were being learned much better than verbs in the Unstag condition but equally well in either staggered condition and equally well in the two *pretraining* tests. It is unclear exactly why this occurred, but it is speculated that verb learning condition required extra attention and working memory that was unavailable in the unstaggered condition due to simultaneous noun and verb learning, but that was available in the StagNoun condition, due to the prior knowledge of nouns, and in the StagVerb condition, due to additional exposure to just verbs.

Overall, the results were compatible with previous XSL studies that suggested syntactic bootstrapping in XSL tasks (Köhne & Crocker, 2010; Alishahi and Pyyukkönen, 2011) and with the emergentist coalition model of word learning (e.g. Hollich et al., 2000; Hirsh-Pasek et al., 2004), which characterizes word learning as resulting from the extraction and interpretation of a variety of different cues. The subsequent section extends the current discussion to the broader topics of a human learning biases and domain-general account of language acquisitions.

CHAPTER FIVE: GENERAL DISCUSSION OF HUMAN LEARNING BIASES AND LINGUISTIC STRUCTURES

This discussion attempts to broaden the idea of multiple cues in word learning to other aspects of language acquisition, and to discuss recent findings in SL research that focus on the mechanisms behind phenomena such as word segmentation, word learning, and syntactic learning (for reviews, see Romberg & Saffran, 2010; Saffran, 2003) in connection with the idea a domain-general account of language acquisition (e.g. Thiesson, 2011).

Recently, the domain-general account of language acquisition has become a popular alternative to the domain-specific account of language acquisition. The domain-specific view of language acquisition argues that the mechanisms behind language acquisition are specific to the language faculty (e.g. Meisel, 1995). According to this theory, language acquisition is constrained by knowledge of specific linguistic structures. A domain-general account of language acquisition, on the other hand, argues that the mechanisms that allow for language architecture are part of a more general human cognitive architecture that include features such as attention, memory, and perception (Colunga & Smith, 2005; Thiesson, 2011). The constraints that arise are a result of these general cognitive features. It is important to note that the domain-general versus domain-specific debate is not one of nature versus nature, but of the mechanism behind the process of language acquisition (Saffran & Thiesson, 2007). Moreover, the acceptance of a domain-general view does not negate the idea of domain-specificity, as several domain-general mechanisms can act together for a domain-specific purpose (Thiesson, 2011).

With regards to word learning, it has been tentatively suggested that a domain-specific idea of constraints and biases (e.g. Markman, 1989) does not adequately capture all aspects of word learning, and that a domain-general account of constraints may provide a better and more comprehensive explanation of word learning (Deák, 2000). Several other accounts of word learning have argued, at least partially, for a domain-general explanation. The XSL account of word learning supports an associative, domain-general mechanism (Yu & Smith, 2007). XSL is classified as a "first-order" statistic (Newport & Aslin, 2004) and is discussed within the field of SL research, which supports a domain-general account of language acquisition due to experiments demonstrating that the same mechanisms used in linguistic domains are also used in nonlinguistic domains and in nonhuman animals (Saffran & Thiesson, 2007). The emergentist coalition model contends that associative and perceptual cues, which are *statistical* cues, are of primary importance in young infants (e.g. Hollich et al., 2000). The emergentist coalition model supports the idea of statistical cues playing a prominent role in young infants and a decreasingly important role in older children (Hollich et al., 2000). Therefore, several of the previously discussed accounts of word learning offer at least partial or indirect support for the domain-generality of the phenomenon.

Sections 5.1 and 5.2 discuss how SL research supports both an integrative view of language acquisition, where statistical, social, and linguistic cues are all important, and a domain-general view of language acquisition.

5.1. HUMAN LEARNING BIASES

SL research has revealed that humans have learning biases that constrain language acquisition. Humans are more sensitive to certain statistical regularities and consequently can more effectively learn certain regularities over others (Saffran, 2002). It is these biases that place constraints on language acquisition, thereby restricting the types of regularities that to which humans are most sensitive (Aslin & Newport, 2008; Endress, Nepor, & Mehler, 2009; Saffran, 2003). In principle, there are numerous statistical regularities in linguistic input that a statistically apt learner could compute, such as the third word in every sentence or the word that follows words whose second syllable begins with *th* (Pinker, 1989). These statistics would be unavailing for humans learning language. Thus, in an environment where an infinite number of statistical computations are possible yet only a select few are actually computed, the argument has been put forth that *humans must be the ideal type of learners of natural language and that they must be able to focus on linguistically relevant statistics and cues* (e.g. Saffran, 2002). Proponents of the SL theory contend that language acquisition is constrained, among other factors, by memory and perception (Aslin & Newport, 2008; Saffran & Thiesson, 2007).

Working memory and long-term memory are obvious constraints on language learning, since language comprehension and production requires knowledge of linguistic patterns. Endress, Carden, Versace, and Hauser (2010) indentify two mechanisms as important for memorizing sequences in linguistic and nonlinguistic domains: tracking co-occurrence frequencies, such as in XSL, and tracking positions of items in sequences.

Auditory perception has also been cited as a crucial component of the cognitive architectural toolbox used for language acquisition. Early accounts of word segmentation seemed to suggest that the *syllable* was the favoured perceptual unit in humans, rather than the *phonemic segment*, because humans seemed particularly good at tracking transitional probabilities between syllables (e.g. Saffran et al., 1996). However, later studies contended against this conclusion. Newport and Aslin (2004) found that adult human learners were poor at learning nonadjacent dependencies when the nonadjacent dependencies were patterned by syllables but not when the nonadjacent dependencies were patterned by consonants or vowels. Newport and Aslin interpreted this as evidence for a strong selection bias for nonadjacent *patterned segments* rather than for syllables.

One idea attempting to explain specific perceptual results of human learners have suggested is that humans are relying on an auditory mechanism similar to the Gestalt principle of similarity (Wertheimer, 1944), which states that some elements are naturally perceived as physically similar or linked together more closely than other elements, even if the other elements are temporally or spatial closer. Another idea regarding constraints on perception, based on work by Endress, Nepor, and Mehler (2009), is that humans have a mechanism dedicated to processing identity relations and that this mechanism is particularly proficient in recognizing identical relations.

The mechanisms behind SL are important for studying linguistic structure and linguistic universals. As previously mentioned, recent evidence has shown that multiple cues, including linguistic and social, are extracted during language acquisition (e.g. Christiansen & Curtin, 2005; Cunillera et al., 2010a; Gerken et al., 2005; Newport & Aslin, 2004; Sahni et al., 2010). A possible question that arises is whether an integrative view of language acquisition *subverts* the domain-general account of language acquisition, given that linguistic or "language-specific" cues (as defined by St. Clair et al., 2010) are also used in language learning. The following subsection argues that an integrative account of language acquisition does not subvert a domain-general account of language acquisition because it is likely that the human learning biases previously discussed have shaped linguistic structures (e.g. Chater & Christiansen, 2010).

5.2. THE EFFECT OF HUMAN LEARNING BIASES ON LINGUISTIC STRUCTURE

Recent work in evolutionary linguistics suggests that many of the properties of language, which are considered to be language-specific (e.g. Sahni et al., 2010), are a result of the human constraints on language learning. In fact, it has been argued that human learning biases have *shaped the structures of natural languages* (e.g. Christiansen & Chater, 2008; Saffran, 2002). This argument suggests that linguistic cues that are thought to be domain-specific are a result of domain-general learning biases.

Computational (e.g. Kirby, 2000; Kirby, 2001) and experimental studies (e.g. Kirby, Smith, & Cornish, 2008) of language transmission and cultural evolution demonstrate how learning biases could shape the structure of natural languages. Languages are culturally transmitted from one generation to the next. The process by which transmission occurs is called iterated learning (Kirby, 1999). Because languages are socially learned and passed on to subsequent generations, it has been argued that languages themselves could be under selection for learnability (e.g. Chater, Reali, & Christiansen, 2009; Smith, 2011; Smith & Kirby, 2008). Languages that are not easily learned will not be learned by future generations and will consequently die out, while languages that are easily learned will continue to exist (Brighton, Kirby, & Smith, 2005). If humans are constrained to detect certain regularities better than others, then a language that is easily learnable will contain these regularities to which humans are more sensitive (Saffran, 2003). Saffran (2003) refers to this as the "constrained statistical learning framework". Therefore, it can be said that the structure of language itself is constrained by human biases (Smith, 2011). This idea could explain the existence of language universals (Chater & Christiansen, 2010; Christiansen & Chater, 2008; Saffran, 2002). If all humans have similar domain-general biases, and if languages are culturally selected for based on easy learnability, then it seems highly possible that all languages, regardless of their surface structure differences, would have the same deep structures (Chater & Christiansen, 2010).

Endress et al. (2009) identify two "perceptual or memory primitives" (henceforth POMPs) that they believe to be two cues that they believe humans are biased to easily identify: reduplication regularities and edge-based positional regularities. They then examined several features of natural language and found that their structures correlated with these two POMPs. For example, they found that reduplication is cross-linguistically a prominent feature of child-directed speech (Ferguson, 1964) and occurs frequently in inflectional morphology, such as in the derivational morphology of Micronesian languages (Moravcsik, 1978). Edge-based positional regularities are also frequently found in natural language. For instance, lexical stress is assigned relative to word edge positions, with stress usually occurring at the left end of a word, such as in Hungarian, or at the right end of a word, such as in French (Hayes, 1995). Furthermore, edge-based positional regularities play a significant role in affixation. Cross-linguistically, prefixes and suffixes are more frequent than infixes, which are affixations that occur in the middle of a word (Broselow & McCarthy, 1983).

In summary, SL researchers and evolutionary linguistics researchers have argued that humans have learning biases that enable them to recognize the pertinent cues in language (Saffran, 2002) and that these learning biases have shaped the structure of natural languages (e.g. Christiansen & Chater, 2008; Saffran, 2002). The evidence presented from computational (e.g. Kirby, 2000; Kirby, 2001) and experimental studies (e.g. Kirby et al., 2008) of language transmission through iterated learning suggests that the cues that seen as linguistic or language-specific (e.g. Sahni et al., 2010) are a result of human learning biases (Christiansen & Chater, 2008; Smith, 2011). According to this research, the fact that certain cues used in language acquisition are language-specific does not contradict a domain-general account of language acquisition. Current ongoing research is attempting to identify the type of human learning constraints, the strength of these constraints, and the effects they have on linguistic structures (e.g. Smith, 2011).

CHAPTER 6: CONCLUSION

Under normal circumstances, children display remarkable word learning abilities (Carey, 1978) even though the input they receive from their environment is full of referential uncertainty (Quine, 1960). There are several different accounts of word learning which focus on different aspects of the process, such as the constraints account (e.g. Markman, 1989); the social-pragmatic account (e.g. Bloom, 2000; Tomasello, 1999); the syntactic and semantic bootstrapping accounts (Gleitman, 1990; Pinker, 1989); and the more recent XSL or associative mechanism approach (e.g. Yu & Smith, 2007). Several authors have suggested that the *combination* of these approaches helps to better explain lexical development (e.g. Frank et al., 2007; Hirsh-Pasek et al., 2000; Smith et al., *under revision*). Hirsh-Pasek and colleagues have introduced the emergentist coalition model of word learning, an integrative model that argues that children extract and interpret multiple interacting cues, including statistical, social, and linguistic (Hirsh-Pasek et al., 2000; Hollich et al., 2000). The scope of this idea can be expanded to include other aspects of language learning, such as word segmentation (Cunillera et al., 2010a) and grammar learning (Frank et al., 2009).

The present work aimed to support the integrative view of word learning presented by the emergentist coalition model (Hirsh-Pasek et al., 2000) through examining the individual and combined contributions of co-occurrence frequencies, a statistical cue, and syntactic-semantic links, a linguistic cue in a sentential XSL experiment. A background literature review explained the individual theories of XSL and syntactic and semantic bootstrapping, and then examined recent experimental (Köhne & Crocker, 2010) and computational (Alishahi & Pyykkönen, 2011) studies that demonstrated how the combination of co-occurrence frequencies and syntactic bootstrapping aids word learning. An original research experiment further examined the combined effect of the cues. In this XSL experiment, participants were instructed to learn novel noun-object and verb-action pairs either with or without the help of syntactic bootstrapping through the use of staggered learning. Results showed that participants scored under 50% when XSL was the primary learning mechanism available, but well over 80% when both XSL and syntactic bootstrapping were available learning mechanisms. The results suggest that bootstrapping is possible and helpful in XSL experiments and that the combination of two different types of cues facilitates word learning.

The general discussion discusses the idea of a domain-general account of language, and whether an integrative account of language acquisitions that draws from statistical, social and linguistic cues, undermines the domain-general account. However, recent studies show that human learning biases may have shaped the structure of languages (Christiansen & Chater, 2010). This indicates that language-specific cues are actually a product of human domain-general learning biases.

In conclusion, emerging evidence suggests that the major questions of language acquisition research, such as how children learn words so quickly and whether language learning is domain-specific or domain general, may be answered if one considers an integrative account of language acquisition where different cues are important at different stages of a child's development. In order to understand the entire scope of this phenomenon, future research should further examine the shift in dependency on specific types of cues based on cognitive development (Hollich et al., 2000), and how human learning biases themselves can create language-specific cues.

REFERENCES

Alishahi, A., & Pyykkönen, P. (2011). The onset of syntactic bootstrapping in word learning: Evidence from a computational study. *Proceedings of the 33nd Annual Conference of the Cognitive Science Society*, Boston.

Au, T. K-F. (1990). Children's use of information in word learning. *Journal of Child Language*. 17, 393-416.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavioral Research Methods, 39(3),* 445–459.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science, 8,* 492 – 499.

Behrend, D. A., Scofield, J., & Kleinknecht, E. E. (2001). Beyond fast mapping: Young children's extensions of novel words and novel facts. *Developmental Psychology, 37(5),* 698-705.

Bloom, P., (1990). Syntactic distinctions in child language. Journal of Child Language 17, 3433 355.

Bloom, P. (1994b). Possible names: The role of syntax-semantics mappings in the acquisition of nominals. *Lingua*, 92, 297-329.

Bloom, P. (2000). *How children learn the meanings of words.* Cambridge, MA: MIT Press.

Broselow, E. and McCarthy, J. (1983) A theory of internal reduplication. *Linguistic Review, 3,* 25–88.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: The MIT Press.

Cattell, R. (2000). *Children's language: consensus and controversy*. London & New York: Continuum International Publishing Group.

Chater, N. & Christiansen, M. H. (2010). Language Acquisition Meets Language Evolution. *Cognitive Science, 34,* 1131-1157.

Chater, N., Reali, F., & Christiansen, M. H. (2009). Restrictions on biological adaptation in language evolution. *Proceedings of the National Academy of Sciences, 106(4),* 1015-1020.

Chomsky, N. 1977. *Reflections on Language.* London: Maurice Temple Smith Ltd.

Christiansen & Kirby. Language evolution: consensus and controversies. *Trends in Cognitive Sciences, 7(7),* 300-307.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31,* 489–558.

Conway & Christiansen. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 24-39.

Cunillera, T., Càmara, E., Laine, M., & Rodríguez-Fornells, A. (2010a). Words as anchors: Known words facilitate statistical learning. *Experimental Psychology, 57,* 134-141.

Cunillera, T., Càmara, E., Laine, M., & Rodríguez-Fornells, A. (2010b). Speech segmentation is facilitated by visual cues. *The Quarterly Journal of Experimental Psychology, 63(2),* 260-274.

Deák, G. (2000). Hunting the Fox of Word Learning: Why "Constraints" Fail to Capture It. Developmental Review, 20, 29-80.

Endress, A.D., Carden, S., Versace, E., & Hauser, M. D. (2010). The apes' edge: positional learning in chimpanzees and humans. *Animal Cognition, 13,* 483-495.

Endress, A. D., Nepor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Science, 13(8),* 348-353.

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, Vol. 59, No. 5, Serial # 242.

Ferguson, C. (1964). Baby talk in six languages. *American Anthropologist, 66(6),* 103-114.

Fiser, J. & Aslin, R.N. (2002). Statistical Learning of Higher-Order Temporal Structure From Visual Shape Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(3),* 458-467.

Fisher, C., Gleitman, H., & Gleitman L. (1991). On the Semantic Content of Subcategorization Frames. *Cognitive Psychology, 23(3),* 331-392.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2007). A Bayesian framework for cross-situational word- learning. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems (volume 20)* (pp. 1212–1222). Cambridge, MA: MIT Press.

Frank M.C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20,* 578- 585.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language, 32,* 249-268.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1 (1),* 3-55.

Hirsh-Pasek, K., Golinkoff, R. M., Hennon, E., & Maguire, M. J. (2004). Hybrid theories at the frontier of developmental psychology: The emergentist coalition model of learning as a case in point. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 173–204). Cambridge, MA: MIT Press.

Hirsh-Pasek, K., Golinkoff, R. M., & Hollich, G. (2000). An emergentist coalition model for word learning: Mapping words to objects is a product of the interaction of multiple cues. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. Smith, A. Woodward, N. Akhtar, M. Tomasello, & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 136-164). New York, NY: Oxford University Press.

Hollich, G. T., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Chung, H. L., Hennon, E., Rocroi, C., & Bloom, L. (2000). Breaking the Language Barrier: An Emergentist Coalition Model for the Origins of Word Learning. *Monographs of the Society for Research in Child Development, 65(3)*, i-vi + 1-135.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month- olds: When speech cues count more than statistics. *Journal of Memory and Language, 44,* 548–567.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics, 61,* 1465–1476.

Kachergis, G., Yu, C., & Shiffrin, R. (2010). Cross-Situational Statistical Learning: Implicit or Intentional? *In S. Ohlsson & R. Catrambone (Eds.), Proceedings of the 32nd Annual Conference of the Cognitive Science Society.* Portland, USA.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In Knight, C., (Ed), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form* (pp. 303-323). Cambridge University Press.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2), 102-110.

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative Cultural Evolution in the Laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105(31),* 10681-10686.

Klein, K., Yu, C., & Shiffrin, R. M. (2008). Prior Knowledge Bootstraps Cross-Situational Learning. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society (pp.715-720).* Austin, Texas.

Köhne, J. & Crocker, M. (2010). "Sentence Processing Mechanisms Influence Cross-Situational Word Learning", in *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society (CogSci),* Portland, USA.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3,* 299-321.

Lidz, J., Gleitman, H., & Gleitman, L. (2004). Kidz in the 'Hood: Syntactic Bootstrapping and the Mental Lexicon. In D. G Hall & S. R. Waxman (Eds). *Weaving a lexicon*. USA: MIT Press.

Macnamara, J. (1982). Names for things: a study of human learning. Cambridge, MA: MIT Press.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11,* F9-F16.

Markman, E. M. (1989). *Categorization and naming in children: problems of induction, Learning, Development, and Conceptual Change*. Cambridge, MA: MIT Press.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*, 91–121.

Maurits, L., Perfors, A. F., & Navarro, D. J. (2009). Joint acquisition of word order and word reference. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society.*

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82,* B101-B111.

Meisel, J. (1995). Parameters in acquisition. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 10–35). Oxford, UK: Blackwell.

Monaghan, P., & Christiansen, M. H. (2008). Integration of multiple probabilistic cues in syntax acquisition. In H. Behrens (Ed.), *Corpora in language acquisition research: History, methods, perspectives* (pp. 139–164). Amsterdam: John Benjamins.

Moravcsik, E. (1978) Reduplicative constructions. In J. H. Greenberg (Ed), *Universals of Human Language: Word Structure (Volume 3),* pp. 297–334. Stanford University Press.

Piccin, T. B. & Waxman, S. R. (2007). Why Nouns Trump Verbs in Word Learning: New Evidence from Children and Adults in the Human Simulation Paradigm. *Language Learning and Development, 3(4),* 295–323.

Pinker, S. (1984) *Language Learnability and Language Development*, Harvard University Press, Cambridge, MA.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.

Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure.* Cambridge, Mass: MIT Press.

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The Birth of Words: Ten-Month-Olds Learn Words Through Perceptual Salience. *Child Development, 77(2),* 266-280.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Romberg, A. R. & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1(6),* 906-914.

Saffran, J. R. (2002). Constraints on Statistical Language Learning. *Journal of Memory and Language, 47, 172*-196.

Saffran, J. R. (2003). Statistical Language Learning: Mechanisms and Constraints. *Current Directions in Psychological Science, 12(4),* 110-114.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928.

Saffran, J.R., & Thiessen, E.D. (2007). Domain-general learning capacities. In E. Hoff & M. Shatz (Eds.), *Handbook of Language Development*. Cambridge: Blackwell (p. 68-86).

Sahni, S. D., Seidenberg, M. S., & Saffran, J. R. (2010). Connecting Cues: Overlapping Regularities Support Cue Discovery in Infancy. *Child Development, 81*, 727-736.

Scott, R. (2010). Learning verbs under referential uncertainty: The role of referential and syntactic contexts. Unpublished doctoral dissertation. Psychology Department, University of Illinois at Urbana-Champaign.

Smith, L.B., Jones, S. & Landau, B. (1996) Naming in young children: A dumb attentional mechanism? *Cognition 60,* 143-171.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106,* 1558-1568.

Smith, K., Smith, A. D. M., & Blythe, R. A. (2010). Cross-Situational Learning: An Experimental Study of Word-Learning Mechanisms. *Cognitive Science, 35(3),* 480-498.

Smith, K., Smith, A. D. M., & Blythe, R. A. (*under revision).* Uncertainty, rapidity and simultaneity: reconsidering human cross-situational learning capacities. *Cognition*.

Snedeker, J., Gleitman, L., & Brent, M. 1999. The Successes and Failures of Word- to-World Mapping. In A. Greenhill, M. Hughs, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the 23rd Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press.

St. Clair, M.C., Monaghan, P., & Christiansen, M. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition, 116,* 341-360.

Thiesson, E., & Saffran, J. R. (2003). When Cues Collide: Use of Stress and Statistical Cues to Word Boundaries by 7- to 9-Month-Old Infants. *Developmental Psychology, 39(4),* 706-716.

Thiesson, E. (2011). Domain General Constraints on Statistical Learning. *Child Development*, *82(2),* 462-470.

Tomasello, M. (1999). *The cultural origins of human cognition*. Boston, MA: Harvard University Press.

Tomasello, M. (2000). The social-pragmatic theory of word learning. *Pragmatics*, *10(4),* 401-413.

Wertheimer, M. (1944). Gestalt theory. *Social Research, 11,* 78–99.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological Review*, *114(2),* 245–272.

Yu, C. & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18(5),* 414-420.

Yurovsky, D., Fricker, D., Yu, C., & Smith, L. B. (2010). "The Active Role of Partial Knowledge in Cross-Situational Word Learning". In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Portland, USA.

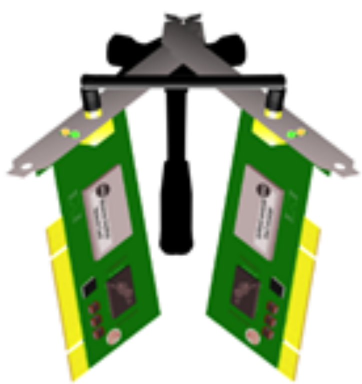
APPENDIX I: EXPERIMENTAL STIMULI

NOUN-OBJECT REFERENT PAIRS

Ablo

Dringer

Foon

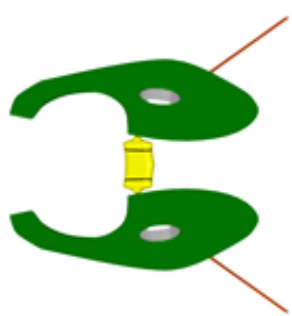


Becime

Feven

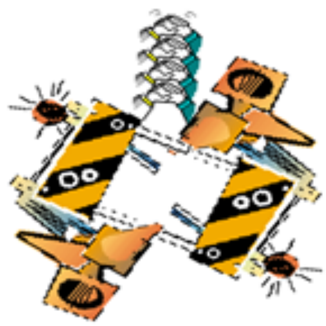
Guba

Cabani

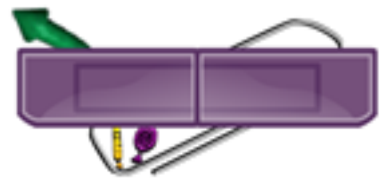


Firch

Himble



Lagonial



Sevy

Mishop

Spoff

Pinen

Tisp

VERB-ACTION PAIRS

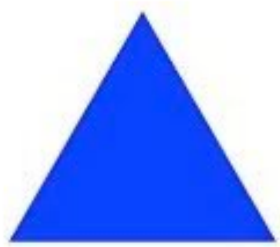


Figure A.I.1: The neutral object used to perform every action during every testing phase and during the pretraining phase in the StagVerb condition.

Verb	**Description of action**
Abisen	Grow, then shrink
Acoption	Flip upside down
Broup	Rotate
Capiga	Zig zag
Garch	Morph
Hidget	Travel in a circle
Hortal	Change colour (to red)
Lugen	Bounce, increasingly smaller
Moga	Flatten
Molax	Move left to right
Noule	Make an X (criss-cross)
Phafe	Jump to all four corners
Seholder	Fade to black
Thevron	Split in half
Vallop	Spiral

Table A.1: The 15 verb-action pairs used in the experiment in Section 2

APPENDIX II: FOIL PROBABILITIES

The following tables indicate all word-referent probabilities based on co-occurrence frequencies in each of the *training* blocks for Version A and B of the experiment described in Section 2. The highlighted numbers indicate the correct word-referent pair.

VERSION A

Word	Referent									
	Ablo	Broup	Cabani	Feven	Garch	Lagonial	Lugen	Noule	Tisp	Vallop
Ablo	1	0.5	0.17	0.33	0.33	0.17	0.5	0.5	0.33	0.33
Broup	0.5	1	0.17	0.5	0	0.5	0.5	0.17	0.33	0.33
Cabani	0.17	0.17	1	0.33	0.67	0.17	0.33	0.33	0.33	0.5
Feven	0.33	0.5	0.17	1	0.17	0.33	0.5	0.17	0	0.33
Garch	0.17	0	0.67	0.33	1	0.17	0.33	0.33	0.33	0.33
Lagonial	0.17	0.5	0.17	0.33	0.17	1	0.33	0.33	0.33	0.67
Lugen	0.5	0.67	0.33	0.5	0.33	0.33	1	0.17	0.33	0.33
Noule	0.67	0.33	0.33	0.17	0.33	0.33	0.17	1	0.5	0.17
Tisp	0.33	0.33	0.17	0	0.33	0.33	0.17	0.67	1	0.5
Vallop	0.17	0.33	0.33	0.33	0.33	0.67	0	0.33	0.33	1

Table A.II.1: Unstaggered condition.

Word	Referent									
	Abisen	Capiga	Dringer	Firch	Guba	Hidget	Mishop	Moga	Seholder	Spoff
Abisen	1	0.17	0.33	0.5	0.5	0.5	0.17	0.17	0.17	0.33
Capiga	0.17	1	0.5	0.17	0.67	0.17	0.17	0.17	0.33	0.5
Dringer	0.33	0.5	1	0.17	0.33	0.33	0.17	0.33	0.33	0.33
Firch	0.5	0.17	0.17	1	0.17	0.67	0.33	0.5	0.17	0.33
Guba	0.5	0.67	0.33	0.17	1	0.33	0.33	0.33	0.17	0.17
Hidget	0.5	0	0.33	0.67	0.33	1	0.33	0.17	0.17	0.17
Mishop	0.17	0	0.33	0.33	0.33	0.17	1	0.67	0.67	0.17
Moga	0.17	0.17	0.33	0.33	0.33	0.17	0.5	1	0.33	0.17
Seholder	0.17	0.33	0.5	0.17	0.17	0.17	0.5	0.33	1	0.33
Spoff	0.5	0.5	0.33	0.17	0.17	0.17	0.17	0.17	0.33	1

Table A.II.2 Staggered-with-nouns (StagNoun) condition.

Word	Referent									
	Acoption	Becime	Casi	Foon	Himble	Hortal	Molax	Pinen	Sevy	Thevron
Acoption	1	0.5	0.33	0	0.5	0	0.17	0.17	0.5	0.5
Becime	0.5	1	0.67	0.17	0.33	0.33	0.17	0.17	0.33	0.17
Casi	0.33	0.67	1	0.33	0.33	0.5	0.17	0.33	0.33	0
Foon	0.17	0.17	0.33	1	0	0.5	0.33	0.33	0.33	0.5
Himble	0.5	0.33	0.33	0	1	0.33	0.33	0.33	0.33	0.5
Hortal	0	0.33	0.5	0.5	0.33	1	0.33	0.5	0.33	0.17
Molax	0.17	0.17	0.17	0.33	0.33	0.33	1	0.67	0.33	0.33
Pinen	0.17	0.17	0.33	0.33	0.33	0.5	0.67	1	0	0.33
Sevy	0.5	0.33	0.33	0.17	0.33	0.33	0.33	0	1	0.5
Thevron	0.5	0.17	0	0.5	0.5	0.17	0.33	0.33	0.5	1

Table A.II.3: Staggered-with-verbs (StagVerb) condition.

VERSION B

Word	Referent									
	Abisen	Dringer	Garch	Guba	Himble	Lugen	Molax	Noule	Spoff	Tisp
Abisen	1	0.17	0.17	0.67	0.33	0.33	0.33	0.17	0.33	0.33
Dringer	0.17	1	0.5	0	0.67	0.33	0.33	0.33	0.17	0.17
Garch	0.17	0.5	1	0.33	0.33	0.5	0	0.33	0.17	0.67
Guba	0.67	0	0.33	1	0.17	0.5	0.17	0.17	0.33	0.5
Himble	0.33	0.67	0.33	0.17	1	0.17	0.5	0.67	0.17	0
Lugen	0.33	0.33	0.5	0.5	0.17	1	0.17	0	0.33	0.5
Molax	0.33	0.5	0	0.17	0.5	0.17	1	0.5	0.67	0.17
Noule	0.17	0.5	0.33	0.17	0.67	0	0.5	1	0.5	0.17
Spoff	0.33	0.17	0.17	0.33	0.17	0.33	0.67	0.5	1	0.33
Tisp	0.33	0.17	0.5	0.5	0	0	0.17	0.17	0.33	1

Table A.II.4: Unstaggered condition.

Word	Referent									
	Ablo	Becime	Broup	Capiga	Casi	Firch	Hidget	Lagonial	Mishop	Moga
Ablo	1	0.33	0.33	0.5	0.5	0	0.5	0.33	0.33	0.17
Becime	0.33	1	0.5	0.5	0.33	0.33	0.33	0.17	0	0.33
Broup	0.33	0.33	1	0.17	0.17	0.5	0.33	0.17	0.33	0.33
Capiga	0.33	0.5	0.17	1	0.33	0.5	0.17	0.33	0.17	0.33
Casi	0.5	0.33	0.17	0.33	1	0.17	0.33	0.5	0.5	0.17
Firch	0	0.5	0.5	0.5	0.17	1	0.33	0.17	0.33	0.5
Hidget	0.5	0.17	0.33	0.17	0.33	0.17	1	0.33	0.5	0.17
Lagonial	0.33	0.17	0.33	0.33	0.5	0.17	0.33	1	0.33	0.5
Mishop	0.33	0	0.33	0.17	0.5	0.33	0.5	0.33	1	0.5
Moga	0.17	0.33	0.33	0.33	0.17	0.5	0.17	0.5	0.5	1

Table A.II.5: Staggered-with-nouns (StagNoun) condition.

Word	**Referent**									
	Acoption	Cabani	Feven	Foon	Hortal	Pinen	Seholder	Sevy	Thevron	Vallop
Acoption	1	0.17	0.17	0.5	0.17	0.67	0	0.5	0.5	0.33
Cabani	0.17	1	0.5	0.17	0.67	0.17	0.33	0	0.33	0.5
Feven	0.17	0.5	1	0.17	0.33	0	0.67	0.33	0.33	0.33
Foon	0.5	0.33	0.17	1	0.33	0.33	0.33	0.17	0.67	0.17
Hortal	0.17	0.67	0.33	0.33	1	0.5	0.33	0.17	0.17	0.33
Pinen	0.67	0.17	0	0.33	0.5	1	0.17	0.5	0.17	0.5
Seholder	0	0.33	0.67	0.33	0.33	0.17	1	0.33	0.33	0.33
Sevy	0.5	0	0.33	0.17	0.17	0.5	0.5	1	0.5	0.33
Thevron	0.5	0.33	0.33	0.67	0.17	0.17	0.33	0.5	1	0
Vallop	0.33	0.33	0.5	0.17	0.33	0.5	0.33	0.33	0	1

Table A.II.6: Staggered-with-verbs (StagVerb) condition.